

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 18:04:28 ; Search time 377 Seconds
(without alignments)
3973.974 Million cell updates/sec

Title: US-09-712-338-2

Perfect score: 2979

Sequence: 1 MRGYEFLSVLPLVAASWALP.....HTQSSVPLPTATSMSSVGMA 555

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cn2_1/USPTO_spool_p/US09712338/runat_17112003_170521_11523/app_query.fasta_1.711
-DB=N_Geneseq 18Jun03 -QFMT=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -NAXLEN=200000000
-USER=US09712338 @CN 1.1 490 @runat_17112003_170521_11523 -NCPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq 19Jun03: *
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT: *
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT: *
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT: *
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT: *
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT: *
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT: *
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT: *
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT: *
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT: *
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT: *
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT: *
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT: *
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT: *
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT: *
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT: *
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT: *
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT: *
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT: *
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT: *
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: *
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT: *
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: *
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: *
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT: *
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2975	99.9	1662	19	AAV28620	A. oryzae ATCC2038
2	2376	79.8	1656	24	ABZ78288	A. niger serine ca
3	2021	67.8	1656	24	ABZ78288	A. niger serine ca
4	980.5	32.9	1872	24	ABZ78289	A. niger serine ca
5	977	32.8	1872	24	ABZ78289	A. niger serine ca
6	944.5	31.7	1665	24	ABZ78283	A. niger serine ca
7	925	31.1	1581	24	ABZ78283	A. niger serine ca
8	902.5	30.3	2940	24	ABZ78184	A. niger serine ca
9	893	30.0	3080	24	ABZ78226	A. niger serine ca
10	704.5	23.6	1007	25	ABZ51997	Aspergillus oryzae
11	636	21.3	2027	24	ABO76315	S. cerevisiae BAX-
12	630	21.1	1611	24	ABZ78243	S. cerevisiae BAX-
13	620	20.8	2002	17	AAI28284	Aspergillus oryzae
14	580	19.5	2068	17	AAI28283	A. niger carboxype
15	580	19.5	2660	24	ABZ78186	A. niger sFAG 2 ca
16	568.5	19.1	2503	15	AAQ55347	A. niger carboxype
17	541.5	18.2	1653	24	ABO76547	Sequence of gene K
18	475	15.9	1551	24	ABZ12878	C. albicans BAX-as
19	460.5	15.5	678	21	AAF12116	Arabidopsis thalia
20	419	14.1	1446	24	ABZ78269	Aspergillus oryzae
21	383	12.9	1510	21	AAI2522	A. niger carboxype
22	375	12.6	626	21	AAF12522	Arabidopsis thalia
23	365	12.3	1368	24	ABZ78267	Aspergillus oryzae
24	361.5	12.1	1936	21	AAI40187	A. niger carboxype
25	361	12.1	1551	19	AAV64076	Arabidopsis thalia
26	361	12.1	1686	22	AAF93808	Human serine carbo
27	359	12.1	1428	22	AAF94477	Human cDNA encodin
28	359	12.1	1633	22	AAF94477	Human hydrophobic
29	359	12.1	1695	21	AAA47444	Human TANGO 176 co
30	359	12.1	1963	22	AAH99751	Human protein enco
31	359	12.1	2076	20	AAH52243	Protein PRO223 cDN
32	359	12.1	2076	22	AAH46129	Human DNA encoding
33	359	12.1	2076	22	AAF72401	Human PRO223 cDNA
34	359	12.1	2076	25	ACA54942	Novel human secret
35	359	12.1	2076	25	ACA54942	Human PRO223 cDNA
36	359	12.1	2076	25	ACA58427	cDNA encoding huma
37	359	12.1	2076	25	ACA60134	Human cDNA for sec
38	359	12.1	2076	25	ABX98357	Human cDNA encodin
39	359	12.1	2076	25	ABX98859	Novel human secret
40	359	12.1	2076	25	ACA05472	cDNA encoding huma
41	359	12.1	2076	25	ACA05904	Human secreted/tra
42	359	12.1	2076	25	ABX96151	Human secreted/tra
43	359	12.1	2076	25	ABX97948	Human PRO polynucl
44	359	12.1	2076	25	ABX78732	Human PRO polynucl
45	359	12.1	2076	25	ABX75745	Human cDNA encodin

ALIGNMENTS

RESULT 1
AAV28620
ID AAV28620 standard; DNA; 1662 BP.
XX
AC AAV28620;
XX
DT 27-AUG-1998 (first entry)
XX
DE A. oryzae ATCC20386 carboxypeptidase I DNA.
XX
KW Carboxypeptidase I; flavour improving agent; hydrolysat; proteinaceous;
XX food industry; ss.
OS Aspergillus oryzae.
XX
FH Key Location/Qualifiers
CDS 1..1662

FT /*tag= a
 FT /product= carboxypeptidase I
 FT /note= "partial coding sequence"

PN W09814599-A1.

PD 09-APR-1998.

PF 03-OCT-1997; 97WO-US17977.

PR 27-NOV-1996; 96US-0757534.

PR 04-OCT-1996; 96US-0726880.

XX (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO-NORDISK AS.

XX Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;

PI Klotz A, Mathisen TE, Rey M;

XX WPI; 1998-240098/21.

DR P-PSDB; AAW56099.

XX Carboxypeptidase from *Aspergillus oryzae* - hydrolyses proteinaceous
 PT substrates, useful for improving flavour of foods

XX Claim 2; Fig 3; 82pp; English.

CC This DNA sequence encodes carboxypeptidase I from *Aspergillus oryzae*.
 CC This polypeptide has an optimal activity in the range of pH 3.0-7.5 at
 CC 25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a
 CC residual activity of at least 65.5% after 30 minutes at pH 4.0 and
 CC 60 deg. C. It also has the capacity to hydrolyse X from N-CBZ-Ala-X where
 CC N-CBZ is N-carboxybenzoyl and X is any amino acid. The carboxypeptidases
 CC can be used for obtaining hydrolysates (which can be enriched in free
 CC glutamic acid or peptide bound glutamic acid residues) from proteinaceous
 CC substrates. The carboxypeptidases can be used in flavour-improving
 CC compositions in the food industry. The products can also be used for the
 CC production of polypeptides free of carboxypeptidase activity.

SQ Sequence 1662 BP; 396 A; 468 C; 398 G; 400 T; 0 other;

Alignment Scores:

Pred. No.: 8,166-276 Length: 1662
 Score: 2975.00 Matches: 554
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatch: 0
 Query Match: 99.87% Indels: 0
 DB: 19 Gaps: 0

US-09-712-338-2 (1-555) x AAV28620 (1-1662)

QY 1 MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuPro 20
 DB 1 ATGCGTGGCTACGAATTTCTCTCAGTGTCTACCCCTTGTTCGACGCGAGTTGGGCGCTTCCA 60
 QY 21 GlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGlyValLys 40
 DB 61 GGAAGTACACCGCGGTCGTCGTAGACAGACGCTACCCAAAGAACCCCGGGGTCAAG 120
 QY 41 ThrLeuThrThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal 60
 DB 121 ACTCTTACACCGCAACATGTCACATCCGCTACAGGAACCCCGGGCGAGGGCGTC 180
 QY 61 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer 80
 DB 181 TGGGAGACTACCCCGGGGTGCAATCTCTCTGGATATGTCGACACCTCTCCCGAGTCC 240
 QY 81 HisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
 DB 241 CATACCTTCTCTCTGGTCTTTCGAGGCCAGACATACCCAGAACTGCACCTATCACATTG 300
 QY 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 120

DB 301 TGGTTGAATGGTGGCCCTCGAAGCGATTCTTGTATCGTCTCTTCGAAGAGTTGGCCCT 360
 QY 121 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSer 140
 DB 361 TGGCATGTCAATTGAGTGTGATGATCAACCTCACTCGTGGAAAGAGTCTCC 420
 QY 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
 DB 421 AATTTACTATTCTCTCCAGCCATTGGGAGTCGGCTTTTCATATAGTAGTACGGTGTAT 480
 QY 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
 DB 481 GGGTCCATTAACCTGTAACTGGGGTGTGCAAAATTCGAGCTTTCGAGGAGTTCAGGGC 540
 QY 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla 200
 DB 541 CGGTACCCACCACTTGATGCCACTCTGATCGATACCAATCTTGGCCGAGAGCCGCT 600
 QY 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
 DB 601 TGGGAGATCCCTGCAAGGATCTCTTAGTGAGTACCTAGCTTGGACTCTAGGCTGAGTCT 660
 QY 221 LysAspPheSerLeuThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn 240
 DB 661 AAGACTTCAGTCTATGACGAGAGCTATGGAGGCACTATGGTCTCTGCAATCTTCAAT 720
 QY 241 HisPheTyrGluGlnAsnGluAlaIleAsnGlySerValAsnGlyValGlnLeuAsn 260
 DB 721 CATTTTACGAGCAGATGAGAGATTGCCAACGGTAGTGTAAATGGTGTTCAGCTTAAT 780
 QY 261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 280
 DB 781 TTCAACTCTCTGGGAATTTATTAACGGCATCATCGACGAGCGCATCCAGGCCCTTACTAC 840
 QY 281 ProGluPheAlaValAsnAsnThrTyrGlyIleIleLysAlaValAsnGluThrValTyrAsn 300
 DB 841 CTTGTAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGCTTACAAC 900
 QY 301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
 DB 901 TACATGAAGTTTGGCAACCAATGCAATGGTTTGGTGGTGGTGTGTATGATATTCGG 960
 QY 321 GlnThrAsnAspThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys 340
 DB 961 CAGACAAACCGCACCGCATTTAGTCTACGCGCTCTTGGCCGCAAGCCACCAACATGTGC 1020
 QY 341 ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 360
 DB 1021 AGGGACATGTTGAGGGGCCATCTACTACGCTTTTGTGTGGTGGTGTGTATGATATTCGG 1080
 QY 361 HisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSer 380
 DB 1081 CATCCATATGATGACCGCATCTCGCCCAAGTTATTACACAAATTTCTGGCAAGGACTCT 1140
 QY 381 ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 400
 DB 1141 GTCATGGAGCTATCGCGCTCAACATCACTACCCAGCTCCCAATTAATGACGCTACTAC 1200
 QY 401 AlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle 420
 DB 1201 GCTTTCAGCAACACAGCGACCTTGTCTGGCCCACTTCATCGAAGACCTCGAGGAGATC 1260
 QY 421 LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp 440
 DB 1261 CTTGTCTCTCCCGTGGCTGTCTCCCTCATCTATGGCGAGCGCCATTACATCTCAACTGG 1320
 QY 441 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 460
 DB 1321 TTCGGGGGTTCAGGCCGCTTCCCTCGCTGGAACTACTCCCAAGCCGCCAGTTCGAGAGC 1380
 QY 461 AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGlnThrArgGluTyrGly 480
 DB 1381 GCAGGGGTACACGCCCTTGAAAGTCAACGGCGCTGAGTATGGGAAATCTCGCGATATGGT 1440

QY 481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 500
 Db 1441 AATTTCCTTCCTACCTCGGCTCTATGAGCGCCATGAGTCCCATCTACCGCCCATC 1500
 QY 501 AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnIlys 520
 Db 1501 GCCTCCCTGCAATGTTTAAACCGACTATCTTCGGTTGGGATATCGCAGAGGCCAGAG 1560
 QY 521 LysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 540
 Db 1561 AAGATCTGGCCACCTCAAGACCAATGGAACGGCTACAGCTACGCATACACATCGTCC 1620
 QY 541 ValProLeuProThrAlaThrSerMetSerSerValGlyMet 554
 Db 1621 GTGCGCTGCTACGCTACCGCTACCGATGTCCAGTGTGGTATG 1662

RESULT 2

ABZ78288
 ID ABZ78288 standard; cDNA; 1656 BP.
 XX AC ABZ78288;
 XX DT 24-APR-2003 (first entry)
 XX OS A. niger serine carboxypeptidase cDNA #3.
 XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ss.
 XX OS Aspergillus niger.

XX FH Key Location/Qualifiers
 FT CDS 1..1656
 FT FT /*cag= a
 FT FT /EC_number= "3.4.16.6"

XX PN W0200268623-A2.

XX PD 06-SEP-2002.

XX PF 22-FEB-2002; 2002WO-EP01984.

XX PR 23-FEB-2001; 2001EP-0200657.

XX PR 23-FEB-2001; 2001EP-0200658.

XX PR 26-FEB-2001; 2001EP-0200660.

XX PR 26-FEB-2001; 2001EP-0200706.

XX PR 26-FEB-2001; 2001EP-0200707.

XX PR 26-FEB-2001; 2001EP-0200708.

XX PR 26-FEB-2001; 2001EP-0200719.

XX PR 28-MAR-2001; 2001EP-0000075.

XX PR 28-MAR-2001; 2001EP-0000078.

XX PR 28-MAR-2001; 2001EP-0000080.

XX PR 28-MAR-2001; 2001EP-0000087.

XX PR 28-MAR-2001; 2001EP-0000088.

XX PR 21-MAY-2001; 2001EP-0000156.

XX PR 21-MAY-2001; 2001EP-0000159.

XX PR 21-MAY-2001; 2001EP-0000160.

XX PR 21-MAY-2001; 2001EP-0000162.

XX PR 21-MAY-2001; 2001EP-0000165.

XX PR 21-MAY-2001; 2001EP-0000166.

XX PR 21-MAY-2001; 2001EP-0000168.

XX PR 21-JUN-2001; 2001EP-0000240.

XX PR 21-JUN-2001; 2001EP-0000242.

XX PR 21-JUN-2001; 2001EP-0000244.

XX PR 21-JUN-2001; 2001EP-0000246.

XX PR 12-JUL-2001; 2001EP-0000280.

XX PR 30-JUL-2001; 2001EP-0000285.

XX PR 30-JUL-2001; 2001EP-0000323.

XX PR 30-JUL-2001; 2001EP-0000327.

XX PR 02-AUG-2001; 2001EP-0000341.

XX PR 02-AUG-2001; 2001EP-0000342.

PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 22-OCT-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0000464.
 PR 21-DEC-2001; 2001EP-00005117.
 XX XX

(STAM) DSM NV.

XX PI Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Stiebler J, Albang R;

XX DR WPI: 2002-723203/78.

XX DR P-PSDB; ABR38864.

XX PT Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications

XX PS Claim 1; Page 252-253; 394pp; English.

XX CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 CC A. niger proteases of the invention.

XX SQ Sequence 1656 BP; 411 A; 426 C; 409 G; 410 T; 0 other;

Alignment Scores:

Pred. No.: 2,86e-218 Length: 1656
 Score: 2376.00 Matches: 428
 Percent Similarity: 87.96% Conservative: 54
 Best Local Similarity: 78.10% Mismatches: 66
 Query Match: 79.76% Indels: 0
 DB: 24 Gaps: 0

US-09-712-338-2 (1-555) x ABZ78288 (1-1656)

QY 1 MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuPro 20
 Db 1 ATGCGTGGCTCTCGGTTCTGGTCTTTGTGCCCCGTGGCTGACCTAGTTGTGCTATGCC 60
 QY 21 GlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGlyVallys 40
 Db 61 GAGATGAATGGTCACTACGATAGAGGCGAGTACCAAAAGCGCTCCACTGGCTGCAA 120
 QY 41 ThrLeuThrThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal 60
 Db 121 TCGATAAAACCCCAACAATGTCTACTATCAGGTATATAAGAACCAAGCAACCAAGGAATT 180
 QY 61 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer 80

Db 181 TGTGAGACACACCTGGGGTCAAAATCATATCTCCGATATGTCGATCTTTCCGCCAGAGTCG 240
Qy 81 HisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
Db 241 CATATCTTCTTTGTGGTTTTCGAGTCAGCGGTGACCCGGAATGATCCAGTACTG 300
Qy 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 120
Db 301 TGGCTGAATGTTGGCCCTGGAGGATTCCTTGAATGGCTTTTGAAGAGTTGGGTCCG 360
Qy 121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 140
Db 361 TGTCAATCACACACAGAGTACCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 420
Qy 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
Db 421 AATCTCTCTTCTTGTCTCAGCCCTCGGTGGGGTCTCTTACAGTGAACCGAGGCC 480
Qy 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
Db 481 GGGTCTTGAATCCATTTACTGGAGCGGTGAGAACGCCCTCTTGTCTGGAGTTCAGGCT 540
Qy 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAla 200
Db 541 CGATACCCAGTTTATGATGCCACTATCATCGACACGACGATATCGCTGCACGCCAAC 600
Qy 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
Db 601 TGGGAGGTCTTCAGGGCTTCTCAGTGGCTTGTCCAGCTAGATTCGGAAGTCAAGTCC 660
Qy 221 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsn 240
Db 661 AAGGAGTTCAACCTGTGGACAGAGAGTTACGGAGGACACTATGACACGAGTTCITCAAT 720
Qy 241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
Db 721 CATTTCTAGCAAAATTCGAAGATCGCTAGCGGGAGAGTCAATGCGCTCCACTGAAT 780
Qy 261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 280
Db 781 TTTAATCTCCTCGGGATTATCAACGCAATCAITGATGCCGATTCAGGCAGACTACTAC 840
Qy 281 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 300
Db 841 GCAGACTTTCGCCCTTAATAATCATATGGAATCAAAAGCTGTCATGACACAGTGTACAAC 900
Qy 301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
Db 901 TATATGAAGTTCGCCAACACAGATGCCAAATGGATGCCAGATCAGGTTCCTTCGTGTAA 960
Qy 321 GlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys 340
Db 961 TTGACCAATAGAGCTTCGCTTCTGATTTATGATATGATACAGACGAGCAATATGTGC 1020
Qy 341 ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 360
Db 1021 AGGGCAATTCGAAGGGCTTACTACAGTTTGGCGCGCGTGGCTGTATGATATTCGG 1080
Qy 361 HisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSer 380
Db 1081 CACCCCTACATGACCGACCGCGCGCTCTTCTTCTTGTGACTACTCTCAAGAAAGACTCA 1140
Qy 381 ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 400
Db 1141 GTCATGATGCTATCGCGGTGGACATTAATACACCGAGTCCAGCGCGGAAGTATATAT 1200
Qy 401 AlaphedGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle 420
Db 1201 GCATTCAGACAGACCGCGGACTTGTATGCGCGAATTCATTGAGGACCTCGAAGATC 1260
Qy 421 LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp 440
Db 1261 CTCAACTCCCGTACGGGTGCTGTGATCTACGGCGATGCCGACTATATCTGTACTGG 1320

Qy 441 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 460
Db 1321 TTCGGCGGTCAAGGCATCTCACTCGCAGTTAACTACCCCATCGAGCTCAGTTCGTGCA 1380
Qy 461 AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 480
Db 1381 GCGGATACACACCATGACAGTAGATGGGTGCGATACGGTGAGACTTCGCGAGTATGGC 1440
Qy 481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 500
Db 1441 AACTTTCTGTTCCACCGGTATATCAGGCTGGCAGGAGTTCATATCATCAACCGATC 1500
Qy 501 AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGlyGlnLys 520
Db 1501 GCAGCGTTCGAGCTTCTCAACCGTACTTTTATTGGATGGGATATTGCAGCGGTACAAC 1560
Qy 521 LysIleTyrProSerTyrIlyThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 540
Db 1561 CAGATTGGCCCGAATATAGCACCAACGGGACATCGAGGCTACACACAGGAGTCGTT 1620
Qy 541 ValProLeuProThrAlaThrSer 548
Db 1621 GTGCCACTGTCCACGGCGTTCGAGT 1644
RESULT 3
ABZ78231
ID ABZ78231 standard; DNA; 3150 BP.
XX AC ABZ78231;
XX DT 24-APR-2003 (first entry)
XX DE A. niger serine carboxypeptidase gene #3.
KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW gene; ds.
OS Aspergillus niger.
XX PN WO200268623-A2.
XX PD 06-SEP-2002.
XX PF 22-FEB-2002; 2002WO-EF01984.
XX 23-FEB-2001; 2001EP-0200657.
PR 23-FEB-2001; 2001EP-0200658.
PR 23-FEB-2001; 2001EP-0200660.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.
PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-JUN-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.
PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.

QY	61	CysGluThrThrProGlyValIysSerTyrSerGlyTyrValAspThrSerProGluSer	80
DB	755	TGTGAGACAACTCTGGGGTCAAAATCATATCTCGGATATGCGATCTTTCCGACAGTCTG	814
QY	81	HisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu	100
DB	815	CATACATTTCTTTGGTTTTCGAGTCAGCCGGTGACCCGAAATGATCCAGTGACTCTG	874
QY	101	TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu	116
DB	875	TGCGTGAATGGTGGCCCTGGAAACGATTCCTTGATTTGGCTTTTTCGA-AGTGTGGCCAAA	933
QY	117	-----GluLeu	118
DB	934	TATCTCTGACGGAAGAGATAAAATTACGTTGTCATGTTCTGACGCCTTTCAACACAGAGTTG	993
QY	119	GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu	138
DB	994	GGTCCGTGTCCATCATCACACGAGAGTACGAATCAATCATCAATCAGTACTCTCTGGAACGAG	1053
QY	139	ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyVal	151
DB	1054	GTCAACCAATCTCTTTCTTGTCGTACGCCCTCGGTGTGGGTATGGAATATTGCTGCCCT	1113
QY	152	-----GlyPheSerTyrSerAspThr	158
DB	1114	CATACATCTCTGAGTACATGCTTACGGTCTTATCTCGAAGGGTTCTCTACAGTGAAC	1173
QY	158	rValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal	178
DB	1174	CGAGGCCGGGTCTCTGAATCCATTTACTGAGCGCTCGAGAACGCCCTCTTTCTCTGGAGT	1233
QY	178	lGlnGlyValGlyProThrIleAspAla	187
DB	1234	TCAGGTGCATACCCAGTTATTGATGCCACTATCATCGGTAGTTCGCGTTTGACTCT	1293
QY	188	-----ThrLeuIleAspThrThrAsnLeuAlaAlaGluAl	199
DB	1294	CACCTAGCATTCCTCCTCAATGCTCTACTTTAGA-GACACGCGCATATCGCTGCACGCGC	1352
QY	199	aAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGln	219
DB	1353	AACCTCGGAGGTGCTTCAGGGCTTCCTCAGTGGCCCTGTCGAGCTAGATTCCGAAGTCAA	1412
QY	219	nSerLysAspPheSerLeuTrpThrGluSerTyrGly-GlyHisTyrGly	235
DB	1413	GTCCAAAGGATTCACTGTGGACAGAGAGTTCAGGAGGGTGAGTGCAACTTTCATACCA	1472
QY	235	-----	235
DB	1473	GACCGACGTAAGTACTTGATCAAGACACTATGGACAGCGGTAGTTGTCTTTCTCTGG	1532
QY	236	-----ProAlaPheAsnHisPheTyrGluGlnAsnGluAl	248
DB	1533	TTGCACACATATTGATCTAATCAGCGAAGTTCTTCAATCATTTCTACGAGCAAAATTCGA	1592
QY	248	rgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleAla	268
DB	1593	AGATCGCTAGCGGGGAAGTCAATGGCGTCCAACTGAATTTTAACTCCCTCGGGATTATCA	1652
QY	268	snGlyIleIleAspGluAlaIleGlnAlaPro	278
DB	1653	ACGGCATCATGATCCCGGATTCAGGTACT-TAGAAATGCAGTTCGCGCAGAGCTGGG	1711
QY	279	-----TyrTyrProGluPheAlaValAla	286
DB	1712	GCTAGAAGGACATCGTAAAGTAATTAATAGGCAGACTACTACGCAGACTTTTGGCGTATA	1771
QY	286	snAsnThrTyrGlyIleIleAla	293
DB	1772	ATPAATCATATGATCAAGCTGTAAAGTTTAAATACACGTACATCGTGGATTTAAGTATC	1831

Qy	294	-----ValAsnGluThrValTyrAsnTyrMetLysPheAlaAsn	306
Db	1832	AAACCGTGCTCATGCTTGGCTAGGTCAATGACACACAGCTGTACAACTATATGAAGTTGCGCAAC	1891
Qy	307	GlnMetProAsnGlyCysGlnAspLeuLeuSerThrCysLysGlnThrAsnArgThrAla	326
Db	1892	ACGATGCCAAATGGATGCCAGGATCAGGTTGCTTCGTGTAAATGTGACCAATAGGACTCG	1951
Qy	327	LeuAlaAspTyrAlaAlaCysAlaGluAlaThrAsnMetCysArgAspAsnVal	344
Db	1952	CTTTCGTATGCTATATGATACAGAGCAGCCAAATATGTGACGGGCAATGT-CGGTGA	2010
Qy	344	-----	344
Db	2011	GTGGTTCTACTGTTTCTCTCAGGGGTGCAATGATGAAGGACTTTCGTAAAGCTGTCTATGT	2070
Qy	345	---GluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr	363
Db	2071	ACGAAGGGGCTTACTACCAAGTTTGGCGCGGTGGCGGTGATGATATTCGGCACCCCTAC	2130
Qy	364	Asp-----	364
Db	2131	AA-TGTAAGTGGCAGAGATBAGGATTGTACTTTCGNAACAGGCACACTGCTCATATGTCA	2189
Qy	365	-----AspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMet	392
Db	2190	ACGTAGGACCGACCCCGCGCTCTACTTTTGTGTGACTACCTCAAGAAGAACTCAGTCATG	2249
Qy	383	AspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPhe	402
Db	2250	GATGCTATCGCGTGGACATTAACTACCGAGTCCAGCGGGAAGATATTATGCAATC	2309
Qy	403	GlnGlnThrGlyAspPheValTyrProAsnPheIleGluAspLeuGluGluIleLeuAla	422
Db	2310	CAGCAGACCGCGACTTTGTATGCGCGAATTCATTGAGGACCTCGNAGAGATCTCCAA	2369
Qy	423	LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGly	442
Db	2370	CTCCCGGTACGCGTGTGTTGATCTACGGGGATGCCGACTATATCTGTAACCTGGTTCGGC	2429
Qy	443	GlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGly	462
Db	2430	GGTCAGGCCAATCTCACTCGCAGTTAACTACCCCATCCAGTCACTAGTCCGTGACGGGGA	2489
Qy	463	TyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPhe	492
Db	2490	TACACCCATCAGACTAGATGGGTGCAATACGGTGAGACTCCGAGATATGCCAACTTT	2549
Qy	483	SerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSer	502
Db	2550	TCGTTCCCCCGGTATATCAGGCTGGGCACGAGGTTCCATACTATCAACCGATCGCAGCG	2609
Qy	503	LeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLysLysIle	522
Db	2610	TTGAGTGTTCACACCGPACTTTTATTGGATGGGATATTCAGCGGGGTACAACCTCAGATT	2669
Qy	523	TrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValPro	542
Db	2670	TGGCCCGNATATAGCACCAACGGGACATCCGACGCTACACACGGAGTCTGTCGTGCCA	2729
Qy	543	LeuProThrAlaThrSer	548
Db	2730	CTGTCCACGGGTGAGT	2747

RESULT 4

ABZ78289

ID ABZ78289 standard; cDNA; 1872 BP.

AC ABZ78289:

XX

DT 24-1

DE
XX

XX	Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW	protein solubility; viscosity; taste; texture; nutritional value;
KW	gene; ss.
XX	
OS	Aspergillus niger.
XX	
XX	
Key	Location/Qualifiers
CDS	1..1872
FT	/*tag= a
FT	/EC_number= "3.4.16.6"
XX	
WN	WO200268623-A2.
XX	
XX	
PD	06-SEP-2002.
XX	
XX	22-FEB-2002; 2002WO-EP01984.
XX	
PR	23-FEB-2001; 2001EP-0200657.
PR	23-FEB-2001; 2001EP-0200658.
PR	23-FEB-2001; 2001EP-0200659.
PR	26-FEB-2001; 2001EP-0200706.
PR	26-FEB-2001; 2001EP-0200707.
PR	26-FEB-2001; 2001EP-0200708.
PR	26-FEB-2001; 2001EP-0200719.
PR	28-MAR-2001; 2001EP-0000075.
PR	28-MAR-2001; 2001EP-0000078.
PR	28-MAR-2001; 2001EP-0000080.
PR	28-MAR-2001; 2001EP-0000087.
PR	28-MAR-2001; 2001EP-0000088.
PR	21-MAY-2001; 2001EP-0000156.
PR	21-MAY-2001; 2001EP-0000159.
PR	21-MAY-2001; 2001EP-0000160.
PR	21-MAY-2001; 2001EP-0000162.
PR	21-MAY-2001; 2001EP-0000165.
PR	21-MAY-2001; 2001EP-0000166.
PR	21-MAY-2001; 2001EP-0000168.
PR	21-JUN-2001; 2001EP-0000240.
PR	21-JUN-2001; 2001EP-0000242.
PR	21-JUN-2001; 2001EP-0000244.
PR	21-JUN-2001; 2001EP-0000246.
PR	12-JUL-2001; 2001EP-0000280.
PR	12-JUL-2001; 2001EP-0000285.
PR	30-JUL-2001; 2001EP-0000323.
PR	30-JUL-2001; 2001EP-0000327.
PR	02-AUG-2001; 2001EP-0000341.
PR	02-AUG-2001; 2001EP-0000342.
PR	02-AUG-2001; 2001EP-0000343.
PR	02-AUG-2001; 2001EP-0000344.
PR	09-AUG-2001; 2001EP-0000357.
PR	16-AUG-2001; 2001EP-0000374.
PR	16-AUG-2001; 2001EP-0000377.
PR	20-SEP-2001; 2001EP-0000478.
PR	20-SEP-2001; 2001EP-0000483.
PR	22-OCT-2001; 2001EP-0000552.
PR	22-OCT-2001; 2001EP-0000553.
PR	22-OCT-2001; 2001EP-0000554.
PR	22-OCT-2001; 2001EP-0000556.
PR	22-OCT-2001; 2001EP-0000557.
PR	22-OCT-2001; 2001EP-0000558.
PR	15-NOV-2001; 2001EP-0004464.
PR	21-DEC-2001; 2001EP-0005117.
XX	
FA	(STAM) DSM NV.
XX	
PI	Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel
PI	Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI	Maier D, Srefioco F, Folkers U, Hopper S, Kemmer W, Tan P;
PI	Stiebler J, Albarg R;
XX	
DR	WPI; 2002-723203/78.
XX	
DR	P-PSDE; ABR38865.
XX	

PT Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
XX applications

PS Claim 1; Page 253-254; 394pp; English.

XX The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in AB278237-AB278293 represent cDNA encoding the
CC A. niger proteases of the invention.

XX Sequence 1872 BP; 402 A; 434 C; 570 G; 466 T; 0 other;

Alignment Scores:

Pred. No.:	3,96e-84	Length:	1872
Score:	980.50	Matches:	216
Percent Similarity:	56.23%	Conservative:	91
Best Local Similarity:	39.56%	Mismatches:	178
Query Match:	32.91%	Indels:	61
DB:	24	Gaps:	19

US-09-712-338-2 (1-555) x AB278289 (1-1872)

QY	44	ThrAlaAsnValThrIleArgTyrLysGlu-----ProGlyAlaGlu-----	59
DB	94	ACGCTGAGATCTCCCGTTATTCATTCGGAGATATTCCTCGTGGAGGATCTCCTAT	153
QY	59	-----GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal	74
DB	154	AMGCAACCCCTCGGCATCTGCACCAACCCCTCCACCCCGAGTACTCTCGGCTAATC	213
QY	75	AspThrSerProGluSer-----	80
DB	214	CACCTCCCGCCACACACCTTACCAATCTCTCCATTCAGGAATCAGCATCTCGCAACCA	273
QY	81	-----HisThrPheThrPhePheGluAlaArg-----HisAsnProGluThr	95
DB	274	TACCCATCAATACCTTTCTGGTACTTTCCTTCCCGCCATCACCACCAATGATACA	333
QY	96	AlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPhe	115
DB	334	TCCCACTCACCATCTGGATGAACGCGCGCGCGGATCTCCATGATGGGCTATT	393
QY	116	GluGluLeuGlyProCysHisValAsnSerThrPheAspAsp---TyrIleAsnProHis	134
DB	394	CAAGAAACCGGCCATGTACTGTGAATACGGACTCGAATTCACGGCCCTATAATCCCTGG	453
QY	135	SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer	154
DB	454	TGTTGGAATGAGTACGCTCGATATGTTGATATTGAGCAGCCGCTGAGACGGGATTTAGT	513
QY	155	TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer	174
DB	514	TATGATGCTGTGAGGAATGGACGTTAGAT-----TTGATGACAGC	555
QY	175	PheAlaGlyValGlnGlyArgTyrProThrIleAsp-----AlaThrLeuIleAspThr	192
DB	556	TTT-----TTGGTGGGGACGTTGCCGAGTCAGGATGTCATGGGACGCTGAATGGGACG	609
QY	193	ThrAsnLeuAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGly-----	210
DB	610	GTTAAT---GGGGGAAGGGCGCTTTGGGTTCGCTTCGAGTTTGGTGGTGAATTTCTCT	666

QY	211	-----LeuProSerLeuAspSerArgValGlnSerLysAsp---PheSerLeuTrpThr	227
DB	667	GAATATGTTTCTTCTGTGACGGGAATGGTGGTGGTATGACAGGGTGAGTATATGACG	726
QY	228	GluSerTyrGlyGlyHisTyrGlyProAlaPheAsnHisPheTyrGluGlnAsnGlu	247
DB	727	GAGTCATATGGGGACGGTATGACCGGCATACACGGCGCTCTTTCAGGAGATGAATGAG	786
QY	248	ArgIleAlaAsnGlySerValAsn---GlyValGlnLeuAsnPheAsnSerLeuGlyIle	266
DB	787	AGGATTGAGAGTGGGAGGTAGACACCGGGAAGATCCATTGGGTACGCTGGGCATT	846
QY	267	IleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsn	286
DB	847	ATCAATGGGGTGTGGATTACTCGTGCAGTCCCTTCGCTCCTGAGCAGGCGTATAAC	906
QY	287	AsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsn	306
DB	907	AATACGATGGGATCGAGGGAATCAATCGCACGCTCTACACCGGCTATGGTAGTTGG	966
QY	307	GlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAla	326
DB	967	AGCAAGCCTGGCGGTGCGAGGATATCATCATCGAGTGTGCGGAG-----1020	
QY	327	LeuAlaAsp---TyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGlu	345
DB	1021	CTCGGAGATCCCTCATCTATCTCGAGGAGCGGTGCGACTACTGTTGCGGGAGATCAAG	1080
QY	346	GlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAsp	365
DB	1081	AGCCTGATACGAATACCTCGGGCGAGGATACGACATAGCGCATTCACGCCGGAT	1140
QY	366	ProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIle	385
DB	1141	GCAGCTCTCTGCTTACTTCGTCGGGTCTTGAATCGCCATGGGTGCAAAAGGCATT	1200
QY	386	GlyValAsnIleAsnThrGlnSerAsnAspValTyrTyrAlaPheGlnGlnThr	405
DB	1201	GGGTCCTCCGCTGAACATATACCATGCTGTCAGAGGCGAGTGGGAAAGTTTCGCTCGACG	1260
QY	406	GlyAspPheValTrpPro-----AsnPheIleGluAspLeuGluIle	420
DB	1261	GGCGATTAT---CCGCGAATGATCCCGCGGATGATCGGGGATATGATATCTG	1314
QY	421	LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp	440
DB	1315	CTTGACTCCGGTCAAGGTGCTATGTTATATGGGACCGGACATATGCTTCTCGGTG	1374
QY	441	PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer	460
DB	1375	CGCGCGGGGAGAGTGTACGCTTGTGGTGGAGTAGAGGATCGCGGAGAGTTCCGTGCT	1434
QY	461	AlaGlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGlu	478
DB	1435	GCTGGGTATGCGAAGTGCAGACGAGTCA---TCCTAGTTGGGGGTCTAGTAAGCGAG	1491
QY	479	TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln	498
DB	1492	TATGGGAATCTTCGTTACGCGTCTTTCAGCGGGCCATGAGTGCCATTTTATCAG	1551
QY	499	ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly	518
DB	1552	CCCGAAACCGGCTATGAGATTTTAAATCGGCTCAGTTTAAATGGGATATTCGACGGGA	1611
QY	519	GlnLysLysIleTrp-----ProSerTyrLysThrAsnGlyThrAlaThrHis	536
DB	1612	GGCATTTCTCTGGACGAGATTCAGAGCTATGGACGGAGGGACCGCTCGTCAACGTGGCAT	1671
QY	537	ThrGlnSerSerValPro	542
DB	1672	ATCAAAAACGAAGTCCCG	1689

RESULT 5

ABZ78232
ID ABZ78232 standard; DNA; 3221 BP.

AC ABZ78232;
XX

DT 24-APR-2003 (first entry)
XX

DE A. niger serine carboxypeptidase gene #4.
XX

XX Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW gene; ds.
XX

OS Aspergillus niger.
XX

XX WO200268623-A2.
XX

PD 06-SEP-2002.
XX

XX 22-FEB-2002; 2002WO-EP01984.
XX

XX 23-FEB-2001; 2001EP-0200657.
XX

XX 23-FEB-2001; 2001EP-0200658.
XX

XX 23-FEB-2001; 2001EP-0200660.
XX

XX 26-FEB-2001; 2001EP-0200706.
XX

XX 26-FEB-2001; 2001EP-0200707.
XX

XX 26-FEB-2001; 2001EP-0200708.
XX

XX 26-FEB-2001; 2001EP-0200719.
XX

XX 28-MAR-2001; 2001EP-0000075.
XX

XX 28-MAR-2001; 2001EP-0000078.
XX

XX 28-MAR-2001; 2001EP-0000080.
XX

XX 28-MAR-2001; 2001EP-0000087.
XX

XX 28-MAR-2001; 2001EP-0000088.
XX

XX 21-MAY-2001; 2001EP-0000156.
XX

XX 21-MAY-2001; 2001EP-0000159.
XX

XX 21-MAY-2001; 2001EP-0000160.
XX

XX 21-MAY-2001; 2001EP-0000162.
XX

XX 21-MAY-2001; 2001EP-0000165.
XX

XX 21-MAY-2001; 2001EP-0000166.
XX

XX 21-MAY-2001; 2001EP-0000168.
XX

XX 21-JUN-2001; 2001EP-0000240.
XX

XX 21-JUN-2001; 2001EP-0000242.
XX

XX 21-JUN-2001; 2001EP-0000244.
XX

XX 21-JUN-2001; 2001EP-0000246.
XX

XX 12-JUL-2001; 2001EP-0000280.
XX

XX 12-JUL-2001; 2001EP-0000285.
XX

XX 30-JUL-2001; 2001EP-0000323.
XX

XX 30-JUL-2001; 2001EP-0000327.
XX

XX 02-AUG-2001; 2001EP-0000341.
XX

XX 02-AUG-2001; 2001EP-0000342.
XX

XX 02-AUG-2001; 2001EP-0000343.
XX

DR

XX P-PSDB; ABR38865.

XX Novel isolated protease polypeptide useful in laboratory, clinical,
XX pharmaceutical, chemical, diagnostic, personal care and industrial
XX applications

PS Claim 1; Page 189-191; 394pp; English.

XX The invention relates to a novel isolated protease polypeptide. A
XX polypeptide or polynucleotide of the invention is useful for diagnosing a
XX fungal infection such as aspergillosis, or as a query sequence to perform
XX a search against public databases. A polypeptide of the invention is
XX useful in a selected number of industrial or pharmaceutical processes, in
XX laboratory or clinical processes, in food industry (baking, brewing,
XX cheese manufacturing, meat tenderising), in tanning industry and in the
XX manufacture of biological detergents. A polypeptide may also be useful
XX for improving protein solubility, extraction yields, viscosity or taste,
XX texture, nutritional value, minimising of antigenicity or
XX anti-nutritional factors, colour or functionality as well as processing
XX aspects like filterability of the proteinaceous raw material. The
XX sequences shown in ABZ78180-ABZ78236 represent genes encoding the
XX A. niger proteases of the invention.

SQ Sequence 3221 BP; 739 A; 773 C; 869 G; 840 T; 0 other;

Alignment Scores:

Pred. No.:	1.84e-83	Length:	3221
Score:	977.00	Matches:	213
Percent Similarity:	55.60%	Conservative:	90
Best Local Similarity:	39.08%	Mismatches:	172
Query Match:	32.80%	Indels:	70
DB:	24	Gaps:	17

US-09-712-338-2 (1-555) x ABZ78232 (1-3221)

QY	59	GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro	78
DB	918	GGCATCTGCACCCACCACCCCTCCACCCCGCTACTCCGGCTACATCACCCTCCCA	977
QY	79	GluSer-----His	81
DB	978	CACACCTTACCAATCTCTCCATTCAGGAATCAGCATCTCGCAACCATCATCAAT	1037
QY	82	ThrPhePheThrPhePheGluAlaArg-----HisAsnProGluThrAlaProIleThr	99
DB	1038	ACCTTTTCTGGTACTTCTTCCTCCGCCATCACCACCAACATCATATCCCACTCAC	1097
QY	100	LeuTyrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGly	119
DB	1098	ATCTGGATGACGGCGCGCGCGGATCTTCATGATTGGGCTATTTCAAGAGACGGG	1157
QY	120	ProCysHisValAsnSerThrPheAspAsp---TyrIleAsnProHisSerTyrAsnGlu	138
DB	1158	CCATGTACTGTGAATACGGACTCGAATTCACGGCCCTATAATCCCTGGTCGTGAATGAG	1217
QY	139	ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr	158
DB	1218	TAGTCGATATGTTGATATTGATGAGCAGCGCGGTGAGACGGGATTTATGATGTTG	1277
QY	159	ValAspGlySerIleAsnProValThr-----	167
DB	1278	AGGAATGGGACGTAGATTGGTTAGTGGGAGATAGATTAGTATTAGTGGTGGAG	1337
QY	168	-----GlyValValGluAsnSerPheAlaGlyValGlnGlyArgTyrProThrIle	185
DB	1338	AGGGATGGAGTAGACAGAATGACGCTTT-----TTGGTGGGACGTGTCGCGAGTCAG	1391
QY	186	Asp-----AlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaLaTyrGluIle	203
DB	1392	GATGTCATGGGACGGTGAATGGACGGTTAAT---GGGGAGAGGGCGCTTTGGTGGCG	1448
QY	204	LeuGlnGlyPheLeuSerGly-----LeuProSerLeuAspSerArgValGln	219

(STAM) DSM NV.
Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel B;
Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
Stiebler J, Albang R;

```

Db      1449  TTGAGGTTTGGTTGGTGAATCTCTCAATATGTTTCTTGTGTGACGGGAATGGTGGT 1508
Qy      220  SerLysAsp---PheSerLeuThrThrClnSerTyrGlyGlyHisTyrGlyProAlaPhe 238
Db      1509  GGTGATGACAGGGTGTAGTATGACGAGGTCATATGGGGACCGGTATGGACCGGCATAC 1568
Qy      239  PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsn---GlyVal 257
Db      1569  ACGCGCTCTTTACGAGAGTAAATGACAGAGATTGAGAGTGGGAGGTAGCACCGGAAG 1628
Qy      258  GlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAla 277
Db      1629  AAGATCCATTGGATACCGTGGCATTTATCAATGGTGTGGATTACTCGTCAGGTC 1698
Qy      278  ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleIleAlaValAsnGluThr 297
Db      1689  CCTTCGTTCCCTGACGAGCGGTATAACAATACGTATGGATCGAGGAAATCAATCGCACG 1748
Qy      298  ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 317
Db      1749  CTCTACGACCGGGCTATGATAGTTGGAGCAAGCTTGGCGGTGCGAGGATATGATCATC 1808
Qy      318  ThrCysLysGlnThr-----AsnArgThrAla 326
Db      1809  GAGTGTGCGGATGTCGCGAGCTCGGAGATCCCTCATGTATGGCGACAATGAGACGTA 1868
Qy      327  LeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGly 346
Db      1869  -----AATGACATCTCGAGGAGCGCTCGGACTCTGTCGCGGAGATCAAGAGC 1919
Qy      347  ProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro 366
Db      1920  CTGTATACGAATACCTCCGCGGAGGATACACACATACGCGATTCACGCCCGGATGCA 1979
Qy      367  ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGly 386
Db      1980  GCTCTCGTGTCTTACTCTGTCGGGTCTTGAATCGCCCATGGTGCAGAAAGGCACATTGG 2039
Qy      387  ValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnGlnThrGly 406
Db      2040  GTCCCGGTGAACATATACCATGTCGTACAGAGCGATGGGGAAACAGTTTCGCTCGAGCGGC 2099
Qy      407  AspPheValTrpPro-----AsnPheIleGluAspLeuGluGluIleLeu 421
Db      2100  GATTAT-----CGCGGAATGATCCCGCGGAATGATCGGGATATGATGACTTGCCT 2153
Qy      422  AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 441
Db      2154  GACTCCGGTGTCAAGGTGGCTATGTATATGGGACCGGACTATGCTTGTCCGTGGCGC 2213
Qy      442  GlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 461
Db      2214  GCGGGGAGAGATGTACGCTGCTGGTGGAGTACAGGATCGGAGAAGTTCGCTGCTGCT 2273
Qy      462  GlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGluTyr 479
Db      2274  GGGTATGCCGAAGTGCAGACGAAGTCA---TCCTACGTTGGGGTCTAGTAGGCGAGTAT 2330
Qy      480  GlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnPro 499
Db      2331  GGGAACTTCTCGTTCACGGGTGCTTTCAGCGGGCCATGAGGTGCCATTTCAGCCCC 2390
Qy      500  IleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGln 519
Db      2391  GAAACGGGTATGAGATTTTAACTCGCTCGTCACTTAATGGGATATTCGACGGGAGGC 2450
Qy      520  LysLysIleTrp-----ProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThr 537
Db      2451  ATTCTCTGGACAGAAATCAGAGCTATGGACGGAGGACCGTCTGTCATCGGTGGCATATC 2510
Qy      538  GlnSerSerValPro 542
      : : : : :

```

```

Db      2511  AAAAACGAAGTGCGC 2525
RESULT 6
ABZ78283
ID      ABZ78283 standard; cDNA; 1665 BP.
XX
AC      ABZ78283;
XX
DT      24-APR-2003 (first entry)
XX
DE      A. niger serine carboxypeptidase cDNA #2.
XX
KW      Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW      protein solubility; viscosity; taste; texture; nutritional value;
KW      gene; ss.
XX
OS      Aspergillus niger.
XX
PH      Key
FT      CDS
FT      1..1665
FT      /*tag= a
FT      /EC_number= "3.4.16.6"
XX
XX      WC200268623-A2.
XX
XX      06-SEP-2002.
XX
XX      22-FEB-2002; 2002WO-EP01984.
XX
XX      23-FEB-2001; 2001EP-0200657.
XX      23-FEB-2001; 2001EP-0200658.
XX      23-FEB-2001; 2001EP-0200660.
XX      26-FEB-2001; 2001EP-0200706.
XX      26-FEB-2001; 2001EP-0200707.
XX      26-FEB-2001; 2001EP-0200708.
XX      26-FEB-2001; 2001EP-0200719.
XX      28-MAR-2001; 2001EP-0000075.
XX      28-MAR-2001; 2001EP-0000078.
XX      28-MAR-2001; 2001EP-0000080.
XX      28-MAR-2001; 2001EP-0000087.
XX      28-MAR-2001; 2001EP-0000088.
XX      21-MAY-2001; 2001EP-0000156.
XX      21-MAY-2001; 2001EP-0000159.
XX      21-MAY-2001; 2001EP-0000160.
XX      21-MAY-2001; 2001EP-0000162.
XX      21-MAY-2001; 2001EP-0000165.
XX      21-MAY-2001; 2001EP-0000166.
XX      21-JUN-2001; 2001EP-0000168.
XX      21-JUN-2001; 2001EP-0000240.
XX      21-JUN-2001; 2001EP-0000242.
XX      21-JUN-2001; 2001EP-0000244.
XX      21-JUN-2001; 2001EP-0000246.
XX      12-JUL-2001; 2001EP-0000280.
XX      12-JUL-2001; 2001EP-0000285.
XX      30-JUL-2001; 2001EP-0000323.
XX      30-JUL-2001; 2001EP-0000327.
XX      02-AUG-2001; 2001EP-0000341.
XX      02-AUG-2001; 2001EP-0000342.
XX      02-AUG-2001; 2001EP-0000343.
XX      02-AUG-2001; 2001EP-0000344.
XX      09-AUG-2001; 2001EP-0000357.
XX      16-AUG-2001; 2001EP-0000374.
XX      16-AUG-2001; 2001EP-0000377.
XX      20-SEP-2001; 2001EP-0000478.
XX      20-SEP-2001; 2001EP-0000483.
XX      22-OCT-2001; 2001EP-000052.
XX      22-OCT-2001; 2001EP-0000553.
XX      22-OCT-2001; 2001EP-0000554.
XX      22-OCT-2001; 2001EP-0000556.
XX      22-OCT-2001; 2001EP-0000557.
XX      15-NOV-2001; 2001EP-0004464.
XX      21-DEC-2001; 2001EP-0005117.

```

XX PA (STAM) DSM NV.

XX PA Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;

XX PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;

XX PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;

XX PI Stiebler J, Albarg R;

XX DR MPI: 2002-723203/78.

XX DR P-PSDB; ABR38859.

XX PT Novel isolated protease polypeptide useful in laboratory, clinical,

XX PT pharmaceutical, chemical, diagnostic, personal care and industrial

XX PT applications

XX PS Claim 1; Page 246-247; 394pp; English.

XX CC The invention relates to a novel isolated protease polypeptide. A

XX CC polypeptide or polynucleotide of the invention is useful for diagnosing a

XX CC fungal infection such as aspergillosis, or as a query sequence to perform

XX CC a search against public databases. A polypeptide of the invention is

XX CC useful in a selected number of industrial or pharmaceutical processes, in

XX CC laboratory or clinical processes, in food industry (baking, brewing,

XX CC cheese manufacturing, meat tenderising), in tanning industry and in the

XX CC manufacture of biological detergents. A polypeptide may also be useful

XX CC for improving protein solubility, extraction yields, viscosity or taste,

XX CC texture, nutritional value, minimising of antigenicity or

XX CC anti-nutritional factors, colour or functionality as well as processing

XX CC aspects like filterability of the proteinaceous raw material. The

XX CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the

XX CC A. niger proteases of the invention.

XX SQ Sequence 1665 BP; 374 A; 455 C; 424 G; 412 T; 0 other;

Alignment Scores:

Pred. No.:	9,66e-81	Length:	1665
Score:	944.50	Matches:	214
Percent Similarity:	51.52%	Conservative:	75
Best Local Similarity:	38.15%	Mismatches:	176
Query Match:	31.71%	Indels:	96
DB:	24	Gaps:	14

US-09-712-338-2 (1-555) x ABZ78283 (1-1665)

QY 14 AlaAlaserTrrAlaLeuProGlySerThrProAlaserValGlyArgArg-GlnLeuPr 33

DB 12 TCTGCTGCTGGGGTCTTCTGGTCTAGGAC-----CGTCAATTTC 56

QY 33 oLysAsnProThrGlyValLysThrLeuThrThra-----AsnAsnValThrleAr 51

DB 57 TCCCGAGCGGAGGCGATCACTGTGCTCAAGTCCAAAGTTGCATGAGAATGTGACTATTTC 116

QY 51 gTyLysGluProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTySe 71

DB 117 TTTCAAGAGCCCT-----GGAATTCGGAACATACCGCGGTGTCGACTATTTC 167

QY 71 rGlyTyrrValAspThrSerProGluSerHisThrPhePheThrPheGluAlaArgHi 91

DB 168 GGGCTATGACACCTTCCCGCCCTCAACAGCTCTTTTGGTTTTCGAAGCGCGAA 227

QY 91 sAsnProGluThrAlaProleThrLeuThrLeuAsnGlyGlyProGlySerAspSerle 111

DB 228 AGATCCCGAGCAATGCGCCTCTGGCCATCTGGCTCAATGGCGGTCCGGGTGCTGCTGCT 287

QY 111 uLleGlyLeuPheGluGluLeuGlyProCysHisValAsnSer---ThrPheAspAspTy 130

DB 288 CARGGGGCTCTTGAAGAATAGTCTCTGTTCCATGTCATCAGACTCCAGACCCACAGT 347

QY 130 rIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGl 150

DB 348 CCTCAATCCTGGAGTGGGAACATGAAGTCAATCTCTATCTCTGACCGGCAACTCA 407

QY 150 yValGlyPheSerTyrrSerAspThrValAspGlySerIleAsnProValThrGlyValva 170

DB 408 AGTCGGCTTCTCATACGATGTCCCAACAAATGGACATTTGACA----- 450

QY 170 lGluAsnSerSerPheAlaGlyValGlnGlyArgTyrrProThrIleAspAlaThrLeuIl 190

DB 451 -----GC 452

QY 190 eAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluLeuLeuGlnGlyPheLeuSerGl 210

DB 453 TAATGGGACTGCATTCGGGCTCAGCTCTATGGCAITTCGGCAAACTGGTTTTCGA 512

QY 210 yLeuProSerLeu-----AspSerArgValGlnSerLysAspPheSerLeuTrpTh 227

DB 513 GTTCCCACTACACAGCAAGCATGATCGTGC-----AGTCTCTGGCG 557

QY 227 rGluSerTyrrGlyHisTyrrGlyProAlaPhePheAsnHisPheTyrrGluGlnAsnGl 247

DB 558 TGAAGTTTACGGAGCCATTATGTCCAGGATCTTTTCGGTTCTTTCCCAACAGCAATGA 617

QY 247 uArgIleAlaAsnGlySerVal---AsnGlyValGln---LeuAsnPheAsnSerLeuGl 265

DB 618 CAAATTCGACAGGAGGACTGCAAGACGGTGCACAGTATTGTCATCTCGACACGCTTGG 677

QY 265 yIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrrTyrrProGluPheAlaVa 285

DB 678 CATTTGTAACGGCTTGATGGATATGCTATCCAGAGAGGCTTACATTACTTGGCCA-- 735

QY 285 lAsnAsnThrTyrrGlyIleLysAlaValAsnGluThrValTyrrAsnTyrrMetLysPheAl 305

DB 736 -----TACAATAACGTAGGCTCGC 755

QY 305 aAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgTh 325

DB 756 CCT-----TCTTCATTCAACTCGCG 776

QY 325 rAlaLeuAlaAspTyrrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp----- 342

DB 777 AGGCTTTCGGATCAGGCCCTCGCTGCGAAGCGGCTTTGAAAGACGCGATTCCGGCTT 836

QY 343 -----AsnVal-----G1 345

DB 837 GCCTCACTCAGGGAAGATATCTCTGAAATTTGCGAGGCGCTTGCACTAGAATGGGAGA 896

QY 345 uGlyPro-----TyrrAlaPheAlaGlyArgGlyValTyrrAspIleArgHisProTy 363

DB 897 TGGCCCATCACCCTACTACACACCTTCATCGGGGTGGTACGATCGCCCTCCTAA 956

QY 363 rAspAspProThrProSerTyrrAsnLysPheLeuAlaLysAspSerValMetAs 383

DB 957 GAACGACCATTCCTCCGCAAGCACATGCTCGGATATTGACGACGAGTCCGCTCTGC 1016

QY 383 pAlaIleGlyValAsnIleAsnTyrrThrGlnSerAsnAsnAspValTyrrAlaPheGl 403

DB 1017 CGCTCTGGGTACCACTGATTCGATTCGAGTATTTGAGTCCGCTGCGCTACACAGTTCT 1076

QY 403 nGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluGluIleLeuAlaLe 423

DB 1077 AAAAACTTTGATTCGTCACGCGGCTTCTCGGATGCAATGGCTACCTCTCTCGACAG 1136

QY 423 uProValArgValSerLeuIleTyrrGlyAspAlaAspTyrrIleCysAsnTrpPheGlyGl 443

DB 1137 TGGTGTAAAGTAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1196

QY 443 yGlnAlaValSerLeuAlaAlaAsnTyrrSerGlnAlaAlaGlnPheArgSerAlaGlyTy 463

DB 1197 CGAAAAAGCCAGCCTTGCAGTTCGCTATTCCTGATTCACCGAATTTGCGACACGGGATA 1256

QY 463 rThrProLeuLysValAsnGlyValGluTyrrGlyGluThrArgGluTyrrGlyAsnPheSe 483

DB 1257 CTCCCACTCTTACGCGCGACGGGATCAGCGGATCAGCGGATCAGCGGATCAGCGGATCAG 1316

QY 483 rPheThrArgValTyrrGluAlaGlyHisGluValProTyrrTyrrGlnProIleAlaSerLe 503

Db 1317 CTTCACTCGCGTCTTCCAAAGCGGGCATGAGGTCCCTCTTACCAGCGTGTGGCGGCTA 1376

QY 503 uGlnLeuPheAsnArgThrIlePheGlyTyrPaspIleAlaGluGlyGlnLysLysIleTr 523

Db 1377 TGAGATCTTATGCGGGGCGCATTCACAAAGATATCCCTACTGGCTCTTGGCTGTGA 1436

QY 523 pProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValProLe 543

Db 1437 TGACGAATCCAGTCGGTGTGACCTAAGGATACGTGGCATATCAAGATATCCCTCTAT 1496

QY 543 u 543

Db 1497 T 1497

RESULT 7

ID ABZ78241 standard; cDNA; 1581 BP.

XX AC ABZ78241;

XX 24-APR-2003 (first entry)

XX A. niger serine carboxypeptidase cDNA #1.

XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;

XX KW protein solubility; viscosity; taste; texture; nutritional value;

XX KW gene; ss.

XX OS Aspergillus niger.

XX PH Key Location/Qualifiers

FT CDS 1..1581

FT FT /*tag= a

FT FT /EC_number= "3.4.16.6"

XX WO200268623-A2.

XX PD 06-SEP-2002.

XX 22-FEB-2002; 2002WO-EP01984.

XX 23-FEB-2001; 2001EP-0200657.

XX 23-FEB-2001; 2001EP-0200658.

XX 23-FEB-2001; 2001EP-0200660.

XX 26-FEB-2001; 2001EP-0200706.

XX 26-FEB-2001; 2001EP-0200707.

XX 26-FEB-2001; 2001EP-0200708.

XX 26-FEB-2001; 2001EP-0200719.

XX 28-MAR-2001; 2001EP-0000075.

XX 28-MAR-2001; 2001EP-0000078.

XX 28-MAR-2001; 2001EP-0000080.

XX 28-MAR-2001; 2001EP-0000087.

XX 28-MAR-2001; 2001EP-0000088.

XX 21-MAY-2001; 2001EP-0000156.

XX 21-MAY-2001; 2001EP-0000159.

XX 21-MAY-2001; 2001EP-0000160.

XX 21-MAY-2001; 2001EP-0000162.

XX 21-MAY-2001; 2001EP-0000165.

XX 21-MAY-2001; 2001EP-0000166.

XX 21-MAY-2001; 2001EP-0000168.

XX 21-JUN-2001; 2001EP-0000240.

XX 21-JUN-2001; 2001EP-0000242.

XX 21-JUN-2001; 2001EP-0000244.

XX 21-JUN-2001; 2001EP-0000246.

XX 12-JUL-2001; 2001EP-0000280.

XX 12-JUL-2001; 2001EP-0000285.

XX 30-JUL-2001; 2001EP-0000323.

XX 02-AUG-2001; 2001EP-0000327.

XX 02-AUG-2001; 2001EP-0000341.

XX 02-AUG-2001; 2001EP-0000342.

XX 02-AUG-2001; 2001EP-0000343.

XX 09-AUG-2001; 2001EP-0000344.

XX 09-AUG-2001; 2001EP-0000357.

PR 16-AUG-2001; 2001EP-0000374.

PR 16-AUG-2001; 2001EP-0000377.

PR 20-SEP-2001; 2001EP-0000478.

PR 20-SEP-2001; 2001EP-0000483.

PR 22-OCT-2001; 2001EP-0000552.

PR 22-OCT-2001; 2001EP-0000553.

PR 22-OCT-2001; 2001EP-0000554.

PR 22-OCT-2001; 2001EP-0000556.

PR 22-OCT-2001; 2001EP-0000557.

PR 22-OCT-2001; 2001EP-0000558.

PR 15-NOV-2001; 2001EP-0000464.

PR 21-DEC-2001; 2001EP-00005117.

XX (STAM) DSM NV.

XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;

XX Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;

XX Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;

XX Stiebler J, Albang R;

XX WPI; 2002-723203/78.

XX P-PSDB; ABR38817.

XX Novel isolated protease polypeptide useful in laboratory, clinical,

XX pharmaceutical, chemical, diagnostic, personal care and industrial

XX applications

XX Claim 1; Page 204-205; 394pp; English.

XX The invention relates to a novel isolated protease polypeptide. A

XX polypeptide or polynucleotide of the invention is useful for diagnosing a

XX fungal infection such as aspergillosis, or as a query sequence to perform

XX a search against public databases. A polypeptide of the invention is

XX useful in a selected number of industrial or pharmaceutical processes, in

XX laboratory or clinical processes, in food industry (baking, brewing,

XX cheese manufacturing, meat tenderising), in tanning industry and in the

XX manufacture of biological detergents. A polypeptide may also be useful

XX for improving protein solubility, extraction yields, viscosity or taste,

XX texture, nutritional value, minimising of antigenicity or

XX anti-nutritional factors, colour or functionality as well as processing

XX aspects like filterability of the proteinaceous raw material. The

XX sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the

XX A. niger proteases of the invention.

XX SQ Sequence 1581 BP; 343 A; 473 C; 423 G; 342 T; 0 other;

Alignment Scores:

Pred. No.:	6.72e-79	Length:	1581
Score:	925.00	Matches:	221
Percent Similarity:	51.46%	Conservative:	78
Best Local Similarity:	38.04%	Mismatches:	192
Query Match:	31.05%	Indels:	90
DB:	24	Gaps:	15

US-09-712-338-2 (1-555) x ABZ78241 (1-1581)

QY 4 TyrGluPheLeuSerValLeuProLeuValAlaAalaSerTrpAlaLeuProGlySerThr 23

Db 4 TACTACTCTCTCTGGTGTCTGCTGCTGGTGGCC-----GCGGTGCC----- 45

QY 24 ProAlaSerValGlyArgGlnLeuProLysAsnProThrGlyValLysThrLeuThr 43

Db 46 -----GTCCTCCGGGCCAGGTTTGTGGCTCCGCCACGAGTCTCATTCGCCACCAAG 96

QY 44 ThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThr 63

Db 97 GGATATCTCGACATCCCGTCCGCTACAAACAGTTC---CCACCGCGATTGTGAGACT 153

QY 64 ThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhe 83

Db 154 GATCCCAAGTGTCAAGAGCTTCTCCGGTTACGTGCGTGTGAGCATGAGCATCTTC 213

QY 84 PheTrpPhePheGluAlaArgHis---AsnProGluThrAlaProIleThrTrpLeu 102

DB 1150 GACTATAAATCGCAATGGCTGGGGAGCCGCTGCTGCTCCAGGTC-----AAG 1200

QY 455 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal---AsnGlyValGluTyr 473

DB 1201 GCGCCAACTTCAGTAGTCGGGTACACCAACATTGTACACCTCGATGGATGACACAC 1260

QY 474 GlyGluThrArgGlyTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGlu 493

DB 1261 GSCCAGTGGCCGACGCGGGCAATTGGCTTTGTGCGAGTGTATCAGAGTGCACATGAG 1320

QY 494 ValProTyrGlnProPheLeuAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyr 513

DB 1321 GTTCCCTTCTATCAACCCCTTGTCTGGCTGGAGATGTTTGACGGCGTCAATGGGSCAAG 1380

QY 514 AspIleAlaGluGlyGlnLysLysIleTyrProSerTyrLysThrAsnGlyThr----- 531

DB 1381 GATGTGGCAGCGGAAGAATTCCCATCTCGTCGAGTTTACACACGGTGGCAGCCCAAG 1440

QY 532 -----AlaThrAlaThr 535

DB 1441 AGTTACTACCGGAGGGCAACAGCACGATTCACTGGGAGGTGTGGATTCTCTGGCGAGC 1500

QY 536 His---ThrGlnSerSerValProLeuProThrAlaThrSerMetSerValGlyMet 554

DB 1501 TACACCAACCAACGAATGCTCCGAACCCGGTGAGCCGAGGCTGAAGCGGATGGGACCA 1560

QY 555 Ala 555

DB 1561 GCT 1563

RESULT 8

ID ABZ78184 standard; DNA; 2940 BP.

XX ABZ78184;

XX 24-APR-2003 (first entry)

XX A. niger serine carboxypeptidase gene #1.

XX Protease; fungal infection; aspergillosis; food; tanning; detergent; protein solubility; viscosity; taste; texture; nutritional value; gene; ds.

XX Aspergillus niger.

XX WO200268623-A2.

XX 06-SEP-2002.

XX 22-FEB-2002; 2002WO-EP01984.

XX 23-FEB-2001; 2001EP-0200657.

XX 23-FEB-2001; 2001EP-0200658.

XX 23-FEB-2001; 2001EP-0200660.

XX 26-FEB-2001; 2001EP-0200706.

XX 26-FEB-2001; 2001EP-0200707.

XX 26-FEB-2001; 2001EP-0200708.

XX 26-FEB-2001; 2001EP-0200719.

XX 28-MAR-2001; 2001EP-0000075.

XX 28-MAR-2001; 2001EP-0000078.

XX 28-MAR-2001; 2001EP-0000080.

XX 28-MAR-2001; 2001EP-0000087.

XX 28-MAR-2001; 2001EP-0000088.

XX 21-MAY-2001; 2001EP-0000156.

XX 21-MAY-2001; 2001EP-0000159.

XX 21-MAY-2001; 2001EP-0000160.

XX 21-MAY-2001; 2001EP-0000162.

XX 21-MAY-2001; 2001EP-0000165.

XX 21-MAY-2001; 2001EP-0000186.

XX 21-JUN-2001; 2001EP-0000188.

XX 21-JUN-2001; 2001EP-0000240.

XX 21-JUN-2001; 2001EP-0000242.

PR 21-JUN-2001; 2001EP-0000244.
 PR 21-JUN-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 12-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000323.
 PR 30-JUL-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 15-NOV-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0000464.
 PR 21-DEC-2001; 2001EP-0005117.
 PA (STAM) DSM NV.
 XX
 XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 PI Klugbauer S, Wagner C, Fritz A, Von Gusted W, Heinrich O;
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmner W, Tan P;
 PI Stiebler J, Albang R;
 XX
 DR WFI; 2002-732203/78.
 DR P-PSDB; ABR38817.
 XX
 XX Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications
 XX
 PS Claim 1; Page 109-111; 394pp; English.
 XX
 CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderizing), in tanning industry and in the
 CC manufacture of biological detergents. In tanning industry and in the
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC anti-nutritional value, minimising of antigenicity or
 CC texture, nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in AB278180-AB278236 represent genes encoding the
 CC A. niger proteases of the invention.
 XX
 SQ Sequence 2940 BP; 691 A; 818 C; 732 G; 699 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,32e-76 Length: 2940
 Score: 902.50 Matches: 231
 Percent Similarity: 49.45% Conservative: 81
 Best Local Similarity: 36.61% Mismatches: 206
 Query Match: 30.30% Indels: 114
 DB: 24 Gaps: 18
 US-09-712-338-2 (1-555) x AB278184 (1-2940)
 QY 4 TyrGluPheLeuSerValLeuProLeuValAlaAlaSerTyrAlaLeuProGlySerThr 23
 Db 456 TACTACTCTCTGGGTGCTGCTGCTGGTGGCC-----GGGTGGCC----- 497
 QY 24 ProAlaSerValGlyArgGlnLeuProLysAsnProThrGlyValLysThrLeuThr 43
 Db 456 TACTACTCTCTGGGTGCTGCTGCTGGTGGCC-----GGGTGGCC----- 497

Db 498 -----GTCTCCCGGGCCAGTTTGTGTGCTCGGCCACGAGATCTCATTTCCCAAG 548
 QY 44 ThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThr 63
 Db 549 GGATATCTCGACATCCCGCTCGCTACAAACAGTGC-----CCACCGGATTTGTGAGACT 605
 QY 64 ThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhe 83
 Db 606 GATCCCAAGTGTCAAGAGCTTCTCGGTGTACGTGCTGCTGAGCATGAGCATCTTC 665
 QY 84 PheTyrPheGluAlaAlaArgHis-----AsnProGluThrAlaProIleThrLeuThrLeu 102
 Db 666 TTCTGGTTCTTCGAGGGCGGCAACAAGATCCCAAGAGGCTCTCTTACCGCTTGGATC 725
 QY 103 AsnGly-----Gly 105
 Db 726 AATGG-AGGCATGTCTGACCCCGGTCAATTAATTTCTTCCATTGCTAACCGTTCTGTAGT 784
 QY 106 ProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyProCysHisValAsnSer 125
 Db 785 CCTGGTTCTCTCCATGATCGGCTTGTTCAGAGCAGCGCCCATCGGATTCAGGCC 844
 QY 126 ThrPheAspTyrIleAsnProHisSerTyrAsnGluValSerAsnLeuLeuPheLeu 145
 Db 845 AATGGCTCCGCTACACAAACCTCTCTCTGGAACAAACGCGCAACATGCTCTCATC 904
 QY 146 SerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnPro 165
 Db 905 GACCAGCGCTGCAGACCGGCTTCTCTCTACAGC-----ATTCCG 943
 QY 166 ValThrGlyValGluAsnSer----- 174
 Db 944 GTTCCCGGTATGTGGATTTCTTCCACAGACAATGTTATTGCTCTGCCCTCCCCGCGCTGC 1003
 QY 175 -----PheAlaGlyValGlnGlyArgTyrProThrIleAsnAla 187
 Db 1004 CCGCACTATGCAGCGGATATGTTCTGTGGCACTTACTTCC-----TACCCCAACGTGAGCCTT 1060
 QY 188 ThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTyrPdpIleLeuGlnGlyPhe 207
 Db 1061 ACGGTAATATCCACCGACAAC---GCGGCCCACTTCTACCGCGCCCTACAGGGTTTT 1117
 QY 208 LeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspSerLeuTyrThr 227
 Db 1118 ATGGGCGCATTTCTCT-----CAGTACTCGCGGAAACCTTCCACTTCCACACG 1165
 QY 228 GluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnGlu 247
 Db 1166 GAGATTATGGCGGCCACHTACGGGCCCGCTTCAACGAGTACATCAGGAGCAGACGCC 1225
 QY 248 ArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIle 267
 Db 1226 CATCTCCAGCGCGGAGCC-----AAGAAGATCAACTGGGCACTGTGATCGGC 1276
 QY 268 AsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsn--- 286
 Db 1277 AATGGCTGTATGACCCCGATTATTCAATACAGGCGCTTACTACACTTACGGTACACTAC 1336
 QY 286 ----- 286
 Db 1337 CTGGTTCTGTATCTTATCTTACAGCTGGACATAATCTATCTACAGGTATATCCGGG 1396
 QY 287 -AsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAs 306
 Db 1397 CAACACATACGACTACCTGCCATTCAACAAGTCCATCAGCTCGCTGATGATGACAAACCT 1456
 QY 306 nGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAl 326
 Db 1457 CTATGCGCCCGAAACTGCTCGACCATCTTACGACTGC-----GCCGC 1501
 QY 326 aLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGl 346
 Db 1502 CCGAGGATCGACGAGATCTGCAGCACTGCCCGCAGATTTTGGCCCAACAGAGGTGAAAA 1561

Qy	346	yProTyrTyralapheAla	gIyVargGlyValTyrAsp	lleAigHisProTyrAsp	Pr 366
Db	1562	CGTCTAGCA	TTACTTCGGTCCGGATAG	TATGACTTCGTGAACT	CACCTCCGACCC 1621
Qy	366	oThrProSerTyrtyrAsn	lyspheLeuAlayspSer	ValMetAspAla	leGl 386
Db	1622	GTTCCTTACGAGT	TTCTACGTTACCTACCT	TGAACAAGCGTCGCT	CAGCGCCATCG 1681
Qy	386	yValAsnIleAsnTyrt	hrClnSerAsnAspVal	TyrTyralapheGln	ThrGl 406
Db	1682	CGCATACATCAAT	TACACGGAGAGCAAC	ACGCTGTGGACTCG	CTTTTTCGTCACCG 1741
Qy	406	y---AspPheValTyr	ProAsnPheIleGlu	AspLeuGluIleLeu	AlaLeuProVa 425
Db	1742	TGACGCGGCGACT	CATGAACACCATCC	GAGGATGTGGCAAG	CTCTCAACAGGGTGT 1801
Qy	425	IArgValSerLeu	IleTyrclyAspAla	AspTyrlleCysAsn	TrpPheGlyGlyClnAl 445
Db	1802	CACGGTGTCTAT	GTACGCCGGGATCCG	CACTATAACTGCA	CTGCTGGTGGGAAGC 1861
Qy	445	aValSerLeuAla	AsnTyrsrGlnAla	AGlnPheArgSerAla	gLyTyrrPr 465
Db	1862	CGTGTCTGTGAGTC	-----AAGCGCGCA	CTTCAGTAGTGGG	TTTACACCAA 1912
Qy	465	oLeuLyVal---	AsnGlyValGluTy	rGlyGluThrArgGlu	TyrrGlyAsnPheSerPh 484
Db	1913	CATTGTCACCTCG	GATGGATGTACACAG	CGCCAGGTGCGC	CAGCGGGCAATTCGCTT 1972
Qy	484	eThrArgValTyrc	luAlaGlyHisGlu	ValProTyrrGlnPro	lleAlaSerLeuGl 504
Db	1973	TGTCGAGTGTATG	AGAGTGGACATAGG	TTCCCTCTATCA	CCCTTGCTGGCTGGA 2032
Qy	504	nLeuPheAsnArg	ThrIlePheGlyTrp	AspIleAlaGluGly	ClnlystIleTrpPr 524
Db	2033	GATGTTTCAGCG	CGTCAATTGGCGCA	AGGATGTGGCAG	CGGAAAGATTCCCATCTCGTC 2092
Qy	524	oSerTyrlYstThr	AsnGlyThr-----	-----	----- 531
Db	2093	GAGTTTACAGAC	GGTGGCGACGCCCA	AGAGTTACTACCG	GAGGGCAACAGCACGATWCA 2152
Qy	532	-----AlaThrAla	ThrHis-----	ThrGlnSerSerVal	ProLeuProTh 545
Db	2153	GTGGAGGTGTGTG	ATCTCTGGCGCAG	TACACACACAC	CAACGATGCTCCGAACCGGT 2212
Qy	545	rAlaThrSerMet	SerSerValGlyMet	Ala 555	
		</			

RESULT 9
ABZ78226

ABZ78226
ID ABZ78226 standard: DNA: 3080 BP.

ID ABZ/8226
XX

XX ABZ78226:

AC ABZ/6ZZ6;
XXXX
DT 24-APR-20DT
XX
24-APR-20

DE A. niger serine carboxypeptidase gene #2.

DE A. HUGER
XX

XX protease; fungal infection; aspergillosis; food; tanning; detergent;
KW

KW protein solubility; viscosity; taste; texture; nutritional value;

gene; ds.

[illegible]

OS Aspergill

XX

PN WO2002686

9
0
6
4
0
6
0
1
0
:
:
X
X

PD 06-SEP-20

XX

PF 22-FEB-2002; 2002WO-EP01984.

XX

PR 23-FEB-20

PR 23-FEB-20

PR	23-FEB-2001	2001EP-0200660
PR	26-FEB-2001	2001EP-0200706
PR	26-FEB-2001	2001EP-0200707
PR	26-FEB-2001	2001EP-0200708
PR	26-FEB-2001	2001EP-0200719
PR	28-MAR-2001	2001EP-0000075
PR	28-MAR-2001	2001EP-0000078
PR	28-MAR-2001	2001EP-0000080
PR	28-MAR-2001	2001EP-0000087
PR	28-MAR-2001	2001EP-0000088
PR	21-MAY-2001	2001EP-0000156
PR	21-MAY-2001	2001EP-0000159
PR	21-MAY-2001	2001EP-0000160
PR	21-MAY-2001	2001EP-0000162
PR	21-MAY-2001	2001EP-0000165
PR	21-MAY-2001	2001EP-0000168
PR	21-JUN-2001	2001EP-0000240
PR	21-JUN-2001	2001EP-0000242
PR	21-JUN-2001	2001EP-0000244
PR	21-JUN-2001	2001EP-0000246
PR	21-JUN-2001	2001EP-0000280
PR	12-JUL-2001	2001EP-0000285
PR	12-JUL-2001	2001EP-0000323
PR	30-JUL-2001	2001EP-0000327
PR	02-AUG-2001	2001EP-0000341
PR	02-AUG-2001	2001EP-0000342
PR	02-AUG-2001	2001EP-0000343
PR	02-AUG-2001	2001EP-0000344
PR	09-AUG-2001	2001EP-0000357
PR	16-AUG-2001	2001EP-0000374
PR	16-AUG-2001	2001EP-0000377
PR	20-SEP-2001	2001EP-0000478
PR	20-SEP-2001	2001EP-0000483
PR	22-OCT-2001	2001EP-0000552
PR	22-OCT-2001	2001EP-0000553
PR	22-OCT-2001	2001EP-0000554
PR	22-OCT-2001	2001EP-0000556
PR	22-OCT-2001	2001EP-0000558
PR	22-OCT-2001	2001EP-0000558
PR	15-NOV-2001	2001EP-0004464
PR	21-DEC-2001	2001EP-0005117

(STAM) ; DSM NV.

Edens L, Van Dijk AA, Krubasik P, Alberman K, Stock A, Kimpel E;
Klugbauer S, Wagner C, Fritz A, Von Gustadt W, Heinrich O;
Mater D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
Stiebler J, Albang R;

WPI: 2002-723203/78.

WPI; 2002-723203/
P-P\$DB: ABR38859.

Novel isolated protease polypeptide useful in laboratory, clinical pharmaceutical, chemical, diagnostic, personal care and industrial applications -

Claim 1: Page 178-180; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A polypeptide or polynucleotide of the invention is useful for diagnosing a fungal infection such as aspergillosis, or as a query sequence to perform a search against public databases. A polypeptide of the invention is useful in a selected number of industrial or pharmaceutical processes, in laboratory or clinical processes, in food industry (baking, brewing, cheese manufacturing, meat tenderising), in tanning industry and in the manufacture of biological detergents. A polypeptide may also be useful for improving protein solubility, extraction yields, viscosity or taste, texture, nutritional value, minimising of antigenicity or anti-nutritional factors, colour or functionality as well as processing aspects like filterability of the proteinaceous raw material. The sequences shown in AB278180-AB278236 represent genes encoding the *A. niger* proteases of the invention.

XX SQ Sequence 3080 BP; 755 A; 803 C; 735 G; 787 T; 0 other;
 Alignment Scores: Pred. No.: 2,03e-75 Length: 3080
 Score: 893.00 Matches: 232
 Percent Similarity: 47.93% Conservative: 81
 Best Local Similarity: 35.53% Mismatches: 200
 Query Match: 29.98% Indels: 141
 DB: 24 Gaps: 19
 US-09-712-338-2 (1-555) x ABZ78226 (1-3080)
 QY 4 TyrGluPheLeuSerValLeuProLeuVal-----AlaAlaSerTrpAlaLeuProGly 21
 Db 490 TATCATTTGCCTTCAGCTTCTTCCCATGTTGAGTAGTCTGCTGCTGGGGTCTTCGGG 549
 QY 22 SerThrProAlaSerValGlyArgArg-GlnLeuProLysAsnProThrGlyValLysTh 41
 Db 550 TCTAGCGAC-----CGCTCAATTTCTCCCGGCGGAGGCATCACTGT 594
 QY 41 rLeuThrThrAla-----AsnAsnValThrIleArgTyrLysGluProGlyAlaGlu-- 58
 Db 595 GCTCAAGTCCAGTTCAGTGCATGAGATGAGCTATTCTTCAAAGAGGTGTGCAGAG-GT 653
 QY 59 -----GlyValCysGluTh 63
 Db 654 ATCTAGAAATAGCTTTTATGCTCGATCGCTGCTGATGTGCACGCTGGAATTTGCGAATC 713
 QY 63 rThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro----- 78
 Db 714 TACGCCGGGTGTCGATCTTATTCGGGTATGTACACCTTCCCGCCGGTTTCTTCTTCGGA 773
 QY 78 ----- 78
 Db 774 CGGACAGGAGAACTGCAGGATATCTCTATCAACAGTAAAGCAATCTCGAATCTGGA 833
 QY 79 -----Glu-SerHisThrPhePheTrpPheGluAlaArgHisAsnP 93
 Db 834 GGATGAGCAATTAAGTGCCTCAACAGCTTCTTTGGTTTTCGAGCCCGCAAGATC 893
 QY 93 roGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleG 113
 Db 894 CCAGCAATCGCTCTGGGCATCTGCTCAATGGCGGTCCGGGTGCTCGTCTCATG 953
 QY 113 lLeuPheGluGluLeuGlyProCysHisValAsnSer---ThrPheAspAspTyrIleA 132
 Db 954 GGCTCTTGAGAAATAGTTCCTTGTTCATTCATGCATCAGACTCCAGACCAAGTCTCTCA 1013
 QY 132 snProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValG 152
 Db 1014 ATCCTTGGAGTTGGAACAATGAAGTCAATCTTATTCCTTGACCAAGCAACTCAAGTCTG 1073
 QY 152 lPheSerTyrSer-----AspThrValAspGlySerIleA 164
 Db 1074 GTTCTTCATACGATGTCACCAAAATGGCATTGTTGGTTCGGACTCGGACGGCGAAGAG 1133
 QY 164 snProValThrGly-----ValValGluAsnSerSerPheAlaGlyValGlnG 180
 Db 1134 AGATAGTTTCCGGTGAATTTCTCCATGTAGTTTCCCGAGTCCCACTTCAACCCATCAFTGTG 1193
 QY 180 lYArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsn-----LeuAlaA 197
 Db 1194 GTACCTTTGCAAGC-----CAGAAGCTTGCACAGACAGCTAATGGAGTCCACACTCAAGCCOA 1307
 QY 197 laGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu----- 214
 Db 1248 CTCACGCTCTAGGCAATTTCCGCAAAACCTGGTTTTTCAGTTCCTCCAGCTGATTCGCGG 1247
 QY 215 ---AspSerArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisT 234
 Db 1308 ACATGATCGTGTG-----AGTCTCTGGCTGAAAGTTTACGGAGGCCATT 1352

234 YrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnGluAArgIleAlaAsnGlySerV 254
 1353 ATGTCAGGCACTTTTCGGTTCTTCCACAGCAGAGATGCAAAATCCGAGGGGACTG 1412
 QY 254 al---AsnGlyValGln---LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleA 272
 Db 1413 CAGAAGAGCGTGCACAGTATTTCATCTCGACAGCTTGGCATTGTGAACGGCTTGATGG 1472
 QY 272 spGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsnAsn----- 287
 Db 1473 ATATGGTATCCAAAGAGGCTTACATTACTTGGCCATCAATAACGTAAGGCTCGCCC 1532
 QY 288 -----ThrTyrGlyIleLysAlaVal 294
 Db 1533 CTTCTTCATTCAACTCGGTAATGCCTAATTCAGTTTCAGACCTACGGCTCGAAATCTTC 1592
 QY 295 AsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAsp 314
 Db 1593 GATAAACCTCTACGAGAACTGATGTATTAACGTGACGCATCCAGGAGGC----- 1643
 QY 315 LeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAla 334
 Db 1644 -----TTTCGGCATCAGGCCCTCGCCTGC 1667
 QY 335 GluAlaThrAsnMetCysArgAsp-----AsnVal----- 344
 Db 1668 GAAGCGCTTTGAAAGACGCGATTCCGGCTTCCTCCTCAGGGAAGATATCTCTGAA 1727
 QY 345 -----GluGlyPro-----TyrTyrAlaPheAla 352
 Db 1728 ATTTCGGAGGCTTCACATAGATGGGAGATGGCCCATCACCCTACCTACACCTTC 1787
 QY 353 GlyArgGlyValTyrAspIleArgHisProTyrAspAspProThrProSerTyrTyr 372
 Db 1788 AATTCGGGTGTGTACGACATCGCCCATCTTAAGAACGACCCATTCCTTCGCAAGCATG 1847
 QY 373 AsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThr 392
 Db 1848 CTCGGATATTGACGAGGAGTCCGCTTCCTGGCGCTCTTGGGTACCACTCAATTTACA 1907
 QY 393 GlnSerAsnAspValTyrTyrAlaPheGlnThrGlyAspPheValTyrProAsn 412
 Db 1908 TCGTCTCTCAGTCCGCTGCTACACAGTTTCAATAAAACCTTTGATATCGTCCAGCGGCG 1967
 QY 413 PheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyrGly 432
 Db 1968 TTCCTGGATGCAATTTGGCTACCTCCTCGACAGTGTGTAAAGTACACATGATGACGA 2027
 QY 433 AspAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyr 452
 Db 2028 GATCGTGATTCGCTGCAATTTGGTTCGGGGCGGAAAGCCAGCCCTTCAGTTCGGTAT 2087
 QY 453 SerGlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGlu 472
 Db 2088 TCCGCTATCACCGAATTTCCGACACGCGATACTCCCACTCTTACGCCCGGAGGATC 2147
 QY 473 TyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHis 492
 Db 2148 AGCGCATGACCGCGCAGCTGGGCAACTACAGCTTCACTCGCTTTCACAGCGGCGCAT 2207
 QY 493 GluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGly 512
 Db 2208 GAGGTCCCTCTACAGCCTGTCCGGCGGTATGAGATCTTCAATCGGCGGACATTTCAAC 2267
 QY 513 TrpAspIleAlaGluGlyGlnLysLysIleTyrProSerTyrLysThrAsnGlyThrAla 532
 Db 2268 AAGATATCCCTACTGGCTCTTGGCTGTGTGATGACGAATTCAGTGGTGGACCTAAG 2327
 QY 533 ThrAlaThrHisThrGlnSerSerValProLeu 543
 Db 2328 GATACGTGCATATCAAGAATATCCCTCTATT 2360

RESULT 10

XX WO200264766-A2.
PN 22-AUG-2002.
XX 21-DEC-2001; 2001WO-EPL5398.
XX 22-DEC-2000; 2000EP-0870318.
PR 04-JAN-2001; 2001EP-0870002.
PR 09-JAN-2001; 2001EP-0870003.
XX (JANC) JANSSEN PHARM NV.
XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
PI WPI; 2002-667002/71.
XX P-PSDB; ABG93049.
DR
XX New isolated nucleic acid representing a synthetic Bax gene, useful as
PT medicament for treating, preventing and/or alleviating yeast or fungal
PT infections or proliferative disorders, or for preventing apoptosis in
PT certain diseases -
XX Claim 36; Figure 1; 344pp; English.
XX This invention describes a novel nucleic acid representing a synthetic
CC Bax gene. The Bax gene of the invention is useful for identifying
CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying
CC Candida spp. sequences that are differentially expressed in a pathway
CC eventually leading to programmed cell death or identifying inhibitors or
CC inhibitor sequences of Bax-induced cell death. The products of the
CC invention have cytostatic, fungicide, immunosuppressive, virucide and
CC vasotropic activity and can be used in vaccines or for gene therapy. The
CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
CC antisense molecules and antibodies are useful as medicaments, or in
CC preparing a medicament for treating, preventing and/or alleviating
CC diseases associated with yeast or fungi or proliferative disorders, such
CC as cancer, or for preventing apoptosis in certain diseases. The compounds
CC or polypeptides, or the genetically modified organism are useful for
CC preparing a medicament for modifying the endogenous flora of humans and
CC other mammals. The vaccine is useful for immunising against yeast or
CC fungal infections. Apoptosis-related diseases include autoimmune disease,
CC ischaemia, diseases related with viral infections or neurodegenerations.
CC This sequence represents a polynucleotide associated with the Bax gene
XX described in the disclosure of the invention.
XX SQ Sequence 2027 BP; 583 A; 380 C; 427 G; 637 T; 0 other;

Alignment Scores:
Pred. NO.: 5.53e-51 Length: 2027
Score: 636.00 Matches: 167
Percent Similarity: 46.29% Conservative: 70
Best Local Similarity: 32.62% Mismatches: 197
Query Match: 21.35% Indels: 78
Db: 24 Gaps: 14

US-09-712-338-2 (1-555) x ABQ076315 (1-2027)

QY 19 LeuProGlySerThrProAlaSerValGlyArgGlnLeu-----ProLys 34
Db 627 CTTCCACAGATACACAGCAACCCCTTAAATGGACCGTTTGAATCAGCATCGCGTG. 686

QY 35 AsnProThrGlyValLysThrLeuThrThrAlaAsnAsnValThrileArgTyrLysGlu 54
Db 687 TTACAACTTTTATCTCTCTGGACACAGATTACAGTTTGAGACTTTAGAACAGTAGAT 746

QY 55 ProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal 74
Db 747 CCTCTCAACTAGGAATT-----GACACCGTAAACAAATGGTCGGGTACATG 794

QY 75 AspThrSerProGluSerHisThrPheThrPhePheGluAlaArgHisAsnProGlu 94
Db 795 GACTATAAGGATTCCAAACACTTTTTTACTGGTTTTTGAAGTAGGAACGATCCTGCT 854

QY 95 ThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 114
Db 855 AACGACCCCAATTATCTTTGGTTAAATGGTGGACCTGGTTGTTCTCGTTTACTGGGTG 914
QY 115 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspSerTyrIleAsnProHis 134
Db 915 CTATTTGAAGTAGGCCCTCATCAATGGCGCGATATGAACCAATCCCAATCCCAT 974
QY 135 SerTrpAsnGluValSerAsnLeuPheLeuSerGlnProLeuGlyValGlyPheSer 154
Db 975 TCTTGAATAATAACCGCTTCAATGATCTCTTGAACAGCACTGGAGTCGGCTTTCC 1034
QY 155 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValGluAsnSerSer 174
Db 1035 TATGGTGAT----- 1043
QY 175 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsn 194
Db 1044 -----GAAAAAGTCTCTCTACAAAA 1064
QY 195 LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu 214
Db 1065 TTAGCAGGCAAGATCGGTACATTTTCTGGATTCTTTTGAAGCTTTCTCATTTA 1124
QY 215 AspSerArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyr 234
Db 1125 -----CGTCCAAACGATTTCCACATTCAGGCGAATCTCTATGCAGACATTAT 1172
QY 235 GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer 253
Db 1173 ATCCCTCAATTCACATGAGATCGTTGTCAGAACCCCTGAAAGA----- 1217
QY 254 ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGlu 273
Db 1218 -----ACGTTCAATTAACTTCAGTTATGTTGTTATGTTATGTTTAAATCAAGT 1412
QY 274 AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys 292
Db 1266 TTGATTCAGCAGATTTATTTATGAACCAATGCGGAAAGGGGGCTATCACCCGTGT 1325
QY 293 AlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 312
Db 1326 CTTCTATCAGAGAATGTGAGAAATGAGTAAGCTGCAGGTCGT----- 1370
QY 313 GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu 332
Db 1371 -----TGTCGTAGTTGAACAAGTTATGTTATGTTCTTAAATCAAGT 1412
QY 333 -----CysAlaGluAlaThrAsnMetCysArgAsnValGluGlyProTyrTyrAla 350
Db 1413 TTACCATGCATAGTCGCCACTGCTTACTGTGACTCTGCACCTTTTGAACCGTACATTAA 1472
QY 351 PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro----- 366
Db 1473 ---ACAGGACTCAACGTCTATGACATTAGAGCGCCCTGTGAAGATAATAGTACTGATGT 1529
QY 367 -----ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 383
Db 1530 ATGTGTTATACAGGTCCTCGCTATGTCGACCAAGTATATGAATTTCTCGTGAAGTTCAAGAA 1589
QY 384 AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPhe 402
Db 1590 ACGTAGGTCGCGAGTCATCAATAATTTATCTGGCTGTGATAATGACGTGTTCCACCGATT 1649
QY 403 GlnGlnThrGlyAspPheValTyrProAsnPheIleGluAspLeuGluIleLeuAla 422
Db 1650 TTGTTTACGGCGATGGGAAGTAACCA---TTTCAACAATATATTTGCTGAATTTAATA 1706
QY 423 LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGly 442
Db 1707 CACAACATTCGGTATTATATATATATGCGGGGTATAGGATTATTTTGAATTTGGCTGGA 1766

QY 443 GlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGly 462
 Db 1767 AACCATGCTTGGTCCATGAGTGAATGGATCAATAAACAGTAGGTATATCAGAGAGGATG 1826
 QY 463 TyrThrPro-----LeuIysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 480
 Db 1827 TTAAGACCATGGTCAGTAAGAAACAGGTGAAGAGTTGGGACAGTCAAGAACTATGGC 1886
 QY 481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProfile 500
 Db 1887 CCTTCACCTTTTCAGATATACGATCCGGTCATATGGTCCCTATGATCAACCGGAG 1946
 QY 501 AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 512
 Db 1947 GCAAGTTTGGAAATGGTCAACAGTTGGATTTCGGGT 1982

RESULT 12

ABZ78243

ID ABZ78243 standard; cDNA; 1611 BP.

XX AC ABZ78243;

XX DT 24-APR-2003 (first entry)

XX DE A. niger carboxypeptidase Y cDNA #1.

XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW Gene; ss.
 XX OS Aspergillus niger.

XX FH Key Location/Qualifiers

XX FT CDS 1..1611

XX FT /*tag= a

XX FT /EC_number= "3.4.16.5"

XX PN WO200268623-A2.

XX PD 06-SEP-2002.

XX PF 22-FEB-2002; 2002WO-EP01984.

XX PR 23-FEB-2001; 2001EP-0200657.

XX PR 23-FEB-2001; 2001EP-0200658.

XX PR 26-FEB-2001; 2001EP-0200660.

XX PR 26-FEB-2001; 2001EP-0200706.

XX PR 26-FEB-2001; 2001EP-0200707.

XX PR 26-FEB-2001; 2001EP-0200708.

XX PR 26-FEB-2001; 2001EP-0200719.

XX PR 28-MAR-2001; 2001EP-0000075.

XX PR 28-MAR-2001; 2001EP-0000078.

XX PR 28-MAR-2001; 2001EP-0000080.

XX PR 28-MAR-2001; 2001EP-0000087.

XX PR 28-MAR-2001; 2001EP-0000088.

XX PR 21-MAY-2001; 2001EP-0000156.

XX PR 21-MAY-2001; 2001EP-0000159.

XX PR 21-MAY-2001; 2001EP-0000160.

XX PR 21-MAY-2001; 2001EP-0000162.

XX PR 21-MAY-2001; 2001EP-0000165.

XX PR 21-MAY-2001; 2001EP-0000166.

XX PR 21-MAY-2001; 2001EP-0000168.

XX PR 21-JUN-2001; 2001EP-0000240.

XX PR 21-JUN-2001; 2001EP-0000242.

XX PR 21-JUN-2001; 2001EP-0000244.

XX PR 21-JUN-2001; 2001EP-0000246.

XX PR 12-JUL-2001; 2001EP-0000280.

XX PR 12-JUL-2001; 2001EP-0000285.

XX PR 30-JUL-2001; 2001EP-0000323.

XX PR 30-JUL-2001; 2001EP-0000327.

XX PR 02-AUG-2001; 2001EP-0000341.

XX PR 02-AUG-2001; 2001EP-0000342.

XX PR 02-AUG-2001; 2001EP-0000343.

PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 22-OCT-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0004464.
 PR 21-DEC-2001; 2001EP-0005117.
 XX (STAM) DSM NV.

XX PI Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Stiebler J, Albang R;
 XX WPI; 2002-723203/78.
 DR P-PSDB; ABR38819.

XX PT Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications -
 XX PS Claim 1; Page 208; 394pp; English.

XX CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 CC A. niger proteases of the invention.

XX SQ Sequence 1611 BP; 326 A; 484 C; 422 G; 379 T; 0 other;

Alignment Scores:

Pred. No.: 1,518-50 Length: 1611
 Score: 630.00 Matches: 167
 Percent Similarity: 47.10% Conservative: 85
 Best Local Similarity: 31.21% Mismatches: 195
 Query Match: 21.15% Indels: 89
 DB: 24 Gaps: 20

US-09-712-338-2 (1-555) x ABZ78243 (1-1611)

QY 7 LeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuProGlySerThrProAlaSer 26
 Db 182 CTCTCTCTCAGCAGGCTCGTAAGCTTTGGGATGAGTGGCCAGCTTCTTC-CCGGAGAGC 240
 QY 27 ValGlyArgArg-----GlnLeuProLys-----AsnProThrGly--- 38
 Db 241 ATGGATCAGAACCTCTCTTTCCCTCCCAAGAGACACACCGCGTCCGACTCGCAC 300
 QY 39 -----ValLysThrLeuThrAlaAsnAsnValThrIleArgTyrLysGlu 54
 Db 301 TGGGACCACATCGTCGATGGCAAGCTGGAAGCTATGATCTCAGGTCAAGAGACCGAT 360
 QY 55 ProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal 74
 Db 361 CCTGGCTCTCTTGGCATC-----GACCCCGCGGTGAAGCAGTACACCGGTATCTC 411

QY	75	AspThrSerPro---GluSerHisThrPhePheTrpPhePheGluAlaArgHisAsnPro	93
Dd	412	GATGACCAAGGAATGATAGCATTTGGTCTACTGGTCTTGAGHCTCGAATGACCC	471
QY	94	GluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlyserAspSerLeuIleGly	113
Dd	472	GAGAATGATCCCGTGTCTGTGGTGAACGGTGGCCCTGGGTGCTTCCTCACC	531
QY	114	LeuPheGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnPro	133
Dd	532	CCTCTCATGGAGCTTGGCCCTAGCAGCATCAACAAGAAATCCACCGGTCTACA	591
QY	134	HisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPhe	153
Dd	592	TACGCTTGGAACTCCAACGCGTCCGTGATCTCTTCCTGACACGCTGTCAAT	651
QY	154	SerTyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSer	173
Dd	652	TCCACAGTAACCTCGCT---	669
QY	174	SerPheAlaGlyValGlnGlyArgTyrProThrilLeAspAlaThrLeuIleAspThr	193
Dd	670	-----GTACGCCGACACG	681
QY	194	AsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSer	213
Dd	682	GTCGCTGTGGCAAGGACGCTCTATGCCTTGCTTACCCTCTCTTCAAACATTCCC	738
QY	214	LeuaspSerArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHis	233
Dd	739	-----GAGTAGCTTAAGCAGGACTTCCACATTGCGCGTGAATCTTATGCTG	789
QY	234	TyrGlyProAlaPheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySer	253
Dd	790	TATATCCCCGCTTCGCTTCGGAGATCCTGTCTCAAGAAGCGC-	834
QY	254	ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleAlaAspGlu	273
Dd	835	-----AACATCAACTGCAGTCCGTTCTCATTTGSCAACGGTCTCACGACGGA	882
QY	274	AlaIleGlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAla	293
Dd	883	TACACCCAGTACGAGTACTACCGTCCCATGGCCCTCGGTGACGCGGTATCCCA	942
QY	294	ValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGln	313
Dd	943	TTGGACAGAGCTCTCCGAGTCATGAC-----AACGCTCTTCCTCGC---TGCAG	993
QY	314	AspLeulleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCys	333
Dd	994	TCTATGATTGAGCTTGCTCAGTTCCGAGAGCGCT-----TGGGTITGT	1038
QY	334	AlaGluAlaThrAsnMetCysArgAspAsnValduGlyProTyrTyrAlaPheAlaGly	353
Dd	1039	GTCCCGGCCCTCCATFCTACTGTATAACACGCCCTCCTTCCGCTTACCAGCGC--	1095
QY	354	ArgGlyValTyrAspIleArgHisProTyrAspAspPro-----ThrPro	368
Dd	1096	CAGNACGTCTATGATGTCCGTGGTAAGTCGAGGATAGCTCTAACCTTGTCTACTCG	1155
QY	369	ProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsn	388
Dd	1156	ATGGGCTACGTACGCGACTACCTGAACAGCCCGAAGTCATCGAGGCTGTGGCGTG	1215
QY	389	IleAsn---TyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAsp	407
Dd	1216	GTCAACGGCTACGACTCGTCAACTTGTACATCAACCGCAACTCTCTTCCACGGTGAC	1275
QY	408	PheValTrpPro-----AsnPheIleGluAspLeuGluGluIleLeuAlaLeuProVal	425
Dd	1276	TGGATGAGCCCTACCAACCGCTCGTTCGGGACTCTCTGAG-----CAGATCCCTGTC	1329

QY	426	ArgValSerLeuIleuTyr--GlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyGln	444
Db	1330	-----TTGAATCTATGCCGTCATGCTCAATTCATTTGCAACTGGCTGGCAACAG	1380
QY	445	AlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrThr	464
Db	1381	GCCTGGACTGAGCCCTGGAGTGGCCGGACAGGCTGAATATGCTCCGCTGAGCTGGAG	1440
QY	465	ProLeuIysVal-----AsnGlyValGluTyrGlyGluThrArgGluTyr	479
Db	1441	GATCTGGTCATTTCGCAATATGACACACACGGCGCAAGAAGATTGGCCAGGTTAAGTCCCAT	1500
QY	480	GlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnPro	499
Db	1501	GGCAACTTCACCTTCATCGCTCTCTATGGTGGTGGCCACATGGTCCCGATGGACACGCC	1560
QY	500	IleAlaSerLeuGlnLeuPheAsnArgThrIlePheGly--Trp	513
Db	1561	GAGTCGAGTCTCGAGTTCCTCACCCGCTGGTGGGAGGTGAATGG	1605
RESULT 13			
AAT28284			
ID	AAT28284 standard; cDNA; 2002 BP.		
XX	AC	AAT28284;	
XX	AC	AAT28284;	
DT	14-AUG-1996 (first entry)		
XX	DT	14-AUG-1996 (first entry)	
DE	A. niger SFAG 2 carboxypeptidase Y cDNA.		
KW	Carboxypeptidase Y; CPY; ascomycete; deuteromycete; host cell;		
KW	protease deficiency; ds.		
XX	Aspergillus niger strain SFAG 2.		
XX	Key	Location/Qualifiers	
FT	CDS	140..1813	
FT	FT	/*tag= a	
FT	mat_peptide	554..1810	
FT	FT	/*tag= b	
XX	XX	WO9609397-A1.	
XX	XX	28-MAR-1996.	
XX	XX	19-SEP-1995; 95WO-US11945.	
XX	XX	20-SEP-1994; 94US-0309341.	
XX	XX	(NOVO) NOVO NORDISK BIOTECH INC.	
XX	XX	Thompson SA, Yaver DS;	
XX	XX	WPI; 1996-188458/19.	
XX	XX	P-PSDB; AAR96738.	
XX	Nucleic acid construct encoding a filamentous ascomycete or		
XX	deuteromycete carboxypeptidase Y - useful to produce host cells		
XX	modified to produce reduced amounts of carboxypeptidase		
XX	Claim 5; Page 23-25; 46pp; English.		
XX	The CPY gene (AAT28284) of Aspergillus niger strain SFAG 2 codes for		
XX	carboxypeptidase Y pre-propeptide (AAR96738). It was obtd. by		
XX	screening a cDNA library in Lambda ZAPII with a probe obtd. by PCR		
XX	amplification of A. niger Bo-1 genomic DNA (see also AAT28283).		
XX	The gene is useful in the creation of CPY-deficient mutants of		
XX	Aspergillus. For example, a selectable marker may be cloned into the		
XX	middle of the CPY gene. The resulting CPY-deficient mutant is useful		
XX	as a host for prodn. of heterologous protein.		
XX	Sequence 2002 BP; 416 A; 530 C; 506 G; 490 T; 0 other.		

Alignment Scores:

Pred. No.: 1,87e-49 Length: 2002
Score: 620.00 Matches: 154
Percent Similarity: 48.25% Conservative: 80
Best Local Similarity: 31.75% Mismatches: 175
Query Match: 20.81% Indels: 76
DB: 17 Gaps: 18

US-09-712-338-2 (1-555) x AAT28284 (1-2002)

```
QY 45 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 64
DB 533 GCCTATGATCTCAGGGTCAAGAAGACCGATCTCTCTGGCATC-----GAC 583
QY 65 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 83
DB 584 CTGGCGTAAGACGATGACCCGGTATCTCGATGACAGGAGACGACGATCTGTC 643
QY 84 PheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsn 103
DB 644 TACTGGTCTTCGAGTCTCGAATGACCCCGAGAATGACCTGTGTTCTGTGGCTGAC 703
QY 104 GlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisVal 123
DB 704 GGTGGCCCTGGATGCTCTCCCTCACCGGTCTTTTCATGGAGCTCGGCCCTAGCAGCATC 763
QY 124 AsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeu 143
DB 764 AACAGAGATCCAGCGGTCTACACGACTAGCTTGAATCCCAACGCGTGGATC 823
QY 144 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 163
DB 824 TTCCTTGACGCGCTGCAACGCGGTACTCTTACAGCAACTCTGCT----- 871
QY 164 AsnProValThrGlyValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 183
DB 871 ----- 871
QY 184 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaTrpGluIle 203
DB 872 -----GTCAGCAGCACCGGTGCTGCTGGCAGGACGCTCATGCGCTTG 913
QY 204 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 223
DB 914 CTTACCTCTCTCTCAACAATTCGCC-----GAGTATGCCAAGCAGGACTTC 961
QY 224 SerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyr 243
DB 962 CACATTGCCGGTGAATCCTATGCTGTCATATATCCCGCTTTGCTCGGAGATTG 1021
QY 244 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 263
DB 1022 TCTCACAGAGAGCGC-----AACATCAACCTCGAGTCC 1054
QY 264 LeuGlyIleLeuAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTyrProGluPhe 283
DB 1055 GTTCTTATTGGCAACGGTCTCACCGCGTCTCAGTACGAGTACTACCGTCCCATG 1114
QY 284 AlaValAsnAsnThrTyrGlyIleLeuAlaValAsnGluThrValTyrAsnTyrMetLys 303
DB 1115 GCCTGTGGTACGGT-----GGTTACCCAGCTGCTTGGACGAG---GGCTCTGCCAGGCC 1168
QY 304 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn 323
DB 1169 ATGGACACACCGCTTCTCTCGC---TGCCAGTCTATGATTGAGTCTTGTCTATAGTTCGAG 1225
QY 324 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 343
DB 1226 AGCGCT-----TGGGTTTGTGTCGCCGCTCCATCTACTGTAAACAACGCC 1270
QY 344 ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr 363
DB 1271 CTCCTTGCCCTTACACGCGC---ACCGACAGAACGCTTACGATGTTGTTGGTAAGTGC 1327
```

```
QY 364 AspAspPro-----ThrProProSerTyrTyrAsnLysPheLeuAlaLys 378
DB 1328 GAGGATAGCTCCAACTCTGCTACTCGGCCAGGCGTACGTCAGGACTACTCTGAACAG 1387
QY 379 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsnAsp 397
DB 1388 ACCGAGGTCATTGAGGCTGTGGCGCTGAGTCAACGGCTAGGACTCGTGCACACTTGAC 1447
QY 398 ValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIleGlu 415
DB 1448 ATCAACCGCAACTTCTCTTCCACGGTGAAGAGCCCTACACCGTCTCTGTTCCG 1507
QY 416 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 434
DB 1508 GGACTCTCTGGAG-----CAGATCCCTGTC-----CTGATCTAGCTGGTACGCC 1552
QY 435 AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGln 454
DB 1553 GATTTCATCTGCAACTGCTGGCAACAAGCGCTGAGTGAAGCCCTTGAGTGGCGCGA 1612
QY 455 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 469
DB 1613 CAGGCTGAATATGCTCCGCTAAGCTGGAGGACCTGGTCTGCTGAGAAATGAGCACAG 1672
QY 470 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 489
DB 1673 GGCAGAGAGATCGGCCAGGTCAAGTCCCATGCACTTCCCTTATGCTGCTCTATGCG 1732
QY 490 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 509
DB 1733 GGTGGCCACATGTCCTCGATGACCAACCCGAGTCGAGTCTTGAATTTCTCAACCGCTGG 1792
QY 510 IlePheGly---Trp 513
DB 1793 TTGGGAGGTGAATGG 1807
RESULT 14
AAT28283
ID AAT28283 standard; DNA; 2068 BP.
XX
AC AAT28283;
XX
DT 14-AUG-1996 (first entry)
XX
DE
XX
XX A. niger Bo-1 genomic carboxypeptidase Y clone.
XX Carboxypeptidase Y; cpy; ascumycete; deuteromycete; host cell;
XX protease deficiency; ds.
XX
OS Aspergillus niger strain Bo-1.
XX
XX Key Location/Qualifiers
FH exon 1..571
FT exon /*tag= a
FT /codon start= 64
FT intron 572..633
FT /*tag= b
FT exon 634..2068
FT /*tag= c
XX
XX WO9609397-AL.
XX
XX 28-MAR-1996.
XX
XX 19-SEP-1995; 95WO-US11945.
XX
XX 20-SEP-1994; 94US-0309341.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX Thompson SA, Yaver DS;
XX
```


Search completed: November 21, 2003, 18:14:22
Job time : 418 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 18:06:28 ; Search time 102 Seconds
(without alignments)

2401.644 Million cell updates/sec

Title: US-09-712-338-2

Perfect score: 2979

Sequence: 1 MRGYEFLSLPLVAASWALP.....HTQSSVPLPLTATSMSSVGMA 555

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Fgapop 6.0			7.0
Delop 6.0			7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool_p/US09712338/runat_17112003_170523_11558/app_query.fasta_1.711
-DB=Issued Patents NA -QPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CGN 1.1 103 @runat_17112003_170523_11558 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database :

```
Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2975	99.9	1662	US-08-943-714-1	Sequence 1, Appli
2	623	20.9	2002	US-08-309-341-3	Sequence 3, Appli
3	623	20.9	2002	US-08-608-267-3	Sequence 3, Appli
4	623	20.9	2002	US-08-608-452-3	Sequence 3, Appli
5	623	20.9	2002	US-08-608-224-3	Sequence 3, Appli
6	623	20.9	2002	US-08-967-149-3	Sequence 3, Appli
7	594	19.9	2068	US-08-309-341-1	Sequence 1, Appli
8	594	19.9	2068	US-08-608-267-1	Sequence 1, Appli
9	594	19.9	2068	US-08-608-452-1	Sequence 1, Appli
10	594	19.9	2068	US-08-608-224-1	Sequence 1, Appli
11	594	19.9	2068	US-08-967-149-1	Sequence 1, Appli
12	568.5	19.1	2503	US-09-640-305-3	Sequence 3, Appli

13	568.5	19.1	2503	1	US-08-360-673-3	Sequence 3, Appli
14	523.5	17.6	2632	2	US-08-899-324-32	Sequence 32, Appli
15	523.5	17.6	2632	3	US-08-329-892B-32	Sequence 32, Appli
16	361	12.1	1551	2	US-08-828-488-4	Sequence 4, Appli
17	361	12.1	1551	4	US-09-299-689A-4	Sequence 4, Appli
18	348.5	11.7	1670	2	US-08-828-488-2	Sequence 2, Appli
19	348.5	11.7	1670	4	US-09-299-689A-2	Sequence 2, Appli
20	339.5	11.4	1814	4	US-09-702-705-319	Sequence 319, App
21	339.5	11.4	1814	4	US-09-736-457-319	Sequence 319, App
22	246.5	8.3	1604	3	US-08-665-966-9	Sequence 9, Appli
23	246.5	8.3	1604	3	US-09-041-780-9	Sequence 9, Appli
24	169.5	5.7	1263	2	US-08-828-488-6	Sequence 6, Appli
25	169.5	5.7	1263	4	US-09-299-689A-6	Sequence 6, Appli
26	163	5.5	1960	4	US-09-280-116-133	Sequence 133, App
27	161.5	5.4	497	4	US-09-480-884A-190	Sequence 190, App
28	161.5	5.4	497	4	US-09-542-615A-190	Sequence 190, App
29	161.5	5.4	497	4	US-09-506-421B-190	Sequence 190, App
30	161.5	5.4	497	4	US-09-280-116-131	Sequence 131, App
31	154	5.2	357	4	US-09-280-116-130	Sequence 130, App
32	154	5.2	357	4	US-08-998-416-160	Sequence 160, App
33	130.5	4.4	522	3	US-09-634-238-44	Sequence 44, Appli
34	127.5	4.3	3279	4	US-09-634-238-44	Sequence 44, Appli
35	127.5	4.3	7818	4	US-08-387-942C-18	Sequence 18, Appli
36	126	4.2	1155	2	US-08-387-942C-1	Sequence 1, Appli
37	126	4.2	12588	4	US-09-865-621A-6	Sequence 6, Appli
38	122.5	4.1	2948	4	US-09-865-621A-8	Sequence 8, Appli
39	122.5	4.1	3427	4	US-09-865-621A-5	Sequence 5, Appli
40	122.5	4.1	3518	4	US-09-865-621A-4	Sequence 4, Appli
41	122.5	4.1	3979	4	US-09-865-621A-3	Sequence 3, Appli
42	122.5	4.1	4810	4	US-09-865-621A-1	Sequence 1, Appli
43	122.5	4.1	5163	4	US-09-252-991A-4847	Sequence 4847, Ap
44	122.5	4.1	5509	4		
45	122	4.1	1962	4		

ALIGNMENTS

RESULT 1

```
US-08-943-714-1
; Sequence 1, Application US/08943714
; Patent No. 6187578
; GENERAL INFORMATION:
; APPLICANT: Binkovsky, Alexander
; APPLICANT: Berk, Randy
; APPLICANT: Rey, Michael
; APPLICANT: Golightly, Elizabeth
; APPLICANT: Klotz, Alan
; APPLICANT: Mathisen, Thomas Erik
; APPLICANT: Dammann, Claus
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids
; TITLE OF INVENTION: Encoding Same
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 61875780 No. 6187578disk of No. 6187578th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,714
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4990.200-US
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...1662
OTHER INFORMATION:
US-08-943-714-1

Alignment Scores:
Pred. No.: 3,44e-308 Length: 1662
Score: 2975.00 Matches: 554
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.87% Indels: 0
DB: 3 Gaps: 0

US-09-712-338-2 (1-555) x US-08-943-714-1 (1-1662)

Qy 1 MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuPro 20
Db 1 ATGCGTGGCTACGAATTTCTCTCACTGCTACCTTGGTGGAGCCAGTTGGGCCCTTCCA 60
Qy 21 GlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGlyValLys 40
Db 61 GGAAGTACACCGGGCTCGGTAGAGACAGACAGTACCCAGAAACCCACCGGGGTCAAG 120
Qy 41 ThrLeuThrThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal 60
Db 121 ACTCTTACAAACCGCAACAAATGTCACTCCGGTACAGAAACCCGGGCGAGGGCGTC 180
Qy 61 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer 80
Db 181 TCGAGAGTACCCCGGGTGTCAATCTCTACTCTGGATATGTGACACCTCTCCCGAGTCC 240
Qy 81 HisThrPheThrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
Db 241 CATACCTTCTCTGGTCTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACATTG 300
Qy 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 120
Db 301 TGGTTGAATGGTGGCCCTGGAGCGGATCTTTGATCGGTCTCTTCGAAAGATTGGGCCCT 360
Qy 121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 140
Db 361 TGCATGTCATTCGACTTTGATGACTATCATCAACCTCTCACTCGTGGAAAGAGGTCTCC 420
Qy 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
Db 421 AATTACTATTCCTGCCAGCCATGGGAGTCGGCTTTTCATATAGTAGATACGGTTGAT 480
Qy 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
Db 481 GGGTCCATTAACTGTTACTGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 540
Qy 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla 200
Db 541 CGGTACCAACCACTGATGCACTCTGATCGATCACTACCAATCTTGGCCGAGAGGGCGCT 600
Qy 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
Db 601 TGGGAGATCTCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTGCAGTCT 660
Qy 221 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsn 240
Db 661 AAGGACTTCAGTCTATGAGCGGAGAGCTATGGAGGGGCACTATGGTCTCTGCTTCTCAAT 720

Qy 241 HisPheTyrGluGlnAsnGlnArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
Db 721 CATTTTACGAGCAGAAATGAGAGAAATGCCAACCGTAGTGTAAATGGTGTTCAGCTTAAT 780
Qy 261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 280
Db 781 TTCAACTCTCTGGGAATTTATACGGCATCATCCAGGAGGATCCAGGCCCTTACTAC 840
Qy 281 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 300
Db 841 CCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGCTCTACAAC 900
Qy 301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
Db 901 TACATGAAGTTTGGCAACCAATGCCAAATGGTTCGCCAGGATTTGATTTCCACCTGCAAA 960
Qy 321 GlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys 340
Db 961 CAGACAAACCGCACCGCATTTAGCTGACTACGCCCTCTCGCCGGAAGCCACCAACATGTGC 1020
Qy 341 ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 360
Db 1021 AGGACAATGTTGAGGGGCCATACCTACGCCCTTGTGCTGTGTGTGTATCATATTCCG 1080
Qy 361 HisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSer 380
Db 1081 CATCCATATGATGACCCGACTCCGCCAAGTATTACAACAATTTCTGCCAAGGACTCT 1140
Qy 381 ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 400
Db 1141 GTCATGGAGCTATCGGCTCAACATCACTACACCCAGTCCCAATATGACCTCTACTAC 1200
Qy 401 AlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle 420
Db 1201 GCTTTCCAGCAACAGGGGACTTTCTGTGCCCAACTTCATCGAAGACCTCGAGGAGATC 1260
Qy 421 LeuAlaLeuProValArgValSerIleuIleTyrGlyAspAlaAspTyrIleCysAsnTyr 440
Db 1261 CTTGTCTCTCCCGTGGTGTCTCCCTCATCTATGGCGAGCGGATTTACATCTGCAACTGG 1320
Qy 441 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 460
Db 1321 TTCGGCGGTACGCGGTTCCTCTCGTGGAACTACTCCAGAGCCGCCAGTTCGGAAGC 1380
Qy 461 AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 480
Db 1381 GCAGGGTACACGCCCTCGAAAGTCAACGCGCTCGAGTATGGGGAACCTCGCGAGTATGTT 1440
Qy 481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 500
Db 1441 AATTTCTCTCTCCTCGCTCTATGAGGCGAGCGCATGAAGTCCCATCTACAGGCCCAATC 1500
Qy 501 AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys 520
Db 1501 GCCTCCCTGCAATGTTTAAACCGGACTATCTTCGGTGGGATATCGCAGAGGCCAGAAAG 1560
Qy 521 LysIleTyrProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 540
Db 1561 AAGATCTGGCCAGCTACAGCAATGGAACGGCTACAGCTACGATACGATACACAGTCGTCC 1620
Qy 541 ValProLeuProThrAlaThrSerMetSerValGlyMet 554
Db 1621 GTCCCGCTCGCTACGGCTACCAAGTCCAGCATGTCAGTGTGGTATG 1662

RESULT 2

US-08-309-341-3
: Sequence 3, Application US/08309341
: Patent No. 5594119
: GENERAL INFORMATION:
: APPLICANT: Yaver, Debbie Sue
: APPLICANT: Thompson, Sheryl Ann
: TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER

824	Db	TTCTTGACGACGCGCTGTCAACGTCGGTACTCTTACACCAACTCTGCT	871
164	Qy	AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro	183
871	Db	-----	871
184	Qy	ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAalaTrpGluIle	203
872	Db	-----GTCAGGACACCGTTCCTCTGCAAGACGAGCTCATGCTT	913
204	Qy	LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe	223
914	Db	CTTACCCTCTTCACAACTCCCC-----GAGTATGCCACGACGAGCTTC	961
224	Qy	SerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyr	243
962	Db	CACATTCCCGTGAATCTATCTGCTGCTCACTATATCCCGCTTTGCTTCGGAGATT	1021
244	Qy	GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer	263
1022	Db	TCTCACAAAGACGCGC-----AACATCAACCTGCAGTCC	1054
264	Qy	LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPhe	283
1055	Db	GTTCATTATGGCAACGCTCACCGACGGTCTCACTCAGTACGAGTACTACCGTCCCATG	1114
284	Qy	AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLys	303
1115	Db	GCCTGTGGTACCGCT--GGTTACCCAGCTGCTTGGACGAG--GGCTCTCCACGCGC	1168
304	Qy	PheAlaasnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn	323
1169	Db	ATGGACAACGCCCTTCTCTCGC--TGCCAGCTCATGATTGAGTCTGCTATAGTTCGAG	1225
324	Qy	ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn	343
1226	Db	AGCGCT-----TGGGTTGTGCCGCGCTTCATCTACTGTATGAACACGCGC	1270
344	Qy	ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr	363
1271	Db	CTCCTTGCCCGCTTACAGCGC--ACCGACACGAGCTCTACCATGTTCTGGTGAAGTGC	1327
364	Qy	AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys	378
1328	Db	GAGGATAGCTCCAACTCTGCTACTCGGCATGGGTACGTCAGCGACTACCTGAACAAG	1387
379	Qy	AspSerValMetAspAlaIleGlyValAsnIleAsn--TyrThrGlnSerAsnAsnAsp	397
1388	Db	ACCGAGGTATTGAGCGCTGTGGCGCTGAGTCAACGGCTACGACTCGTGGCACTTTGAC	1447
398	Qy	ValTyrTyrAlaPheGlnThrGlyAspPheValTyrPro-----AsnPheIleGlu	415
1448	Db	ATCAACCGCAACTTCCTTTCACCGGTGACTGGATGAAGCCCTACACCGCTCTGCTCCG	1507
416	Qy	AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr--GlyAspAla	434
1508	Db	GGACTCTCTGAG-----CAGATCCCTGTC-----CTGATCTACGCTGGTACGCC	1552
435	Qy	AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGln	454
1553	Db	GATTTCATCTGCACTGGCTGGCAACAAGGCGCTTGACTGAAGCCCTTGAGTGGCCCGGA	1612
455	Qy	AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn	469
1613	Db	CAGGCTGAATATGCCTCCCGCTAAGCTGGAGGACCTGGTCTGTCGAGATGAGCAACAG	1672
470	Qy	GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu	489
1673	Db	GGCAAGAAGATCGGCGAGGTCAAGTCCCATGGCACTTCACCTTCATGCTCTCTATGGC	1732
490	Qy	AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr	509

Qy 510 IlePheGly---Tyr 513
 : : : : :
 Db 1793 TTGGGAGTGAATGG 1807

RESULT 3

US-08-608-267-3
 ; Sequence 3, Application US/08608267
 ; Patent No. 5688663

GENERAL INFORMATION:

APPLICANT: Taver, Debbie Sue
 APPLICANT: Thompson, Sheryl Ann
 TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESS: No. 56886630 No. 56886630 disk of No. 5688663th America, Inc.
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/608,267
 FILING DATE: 28-FEB-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/309,341

FILING DATE: 20-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lowney, Karen A.

REGISTRATION NUMBER: 31,274

REFERENCE/DOCKET NUMBER: 4247.000-US

TELEPHONE: 212 867 0123

TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2002 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE:

ORGANISM: Aspergillus niger

FEATURE:

NAME/KEY: intron

LOCATION: 349..411

FEATURE:

NAME/KEY: CDS

LOCATION: join (348..412)

US-08-608-267-3

Alignment Scores:

Pred. No.: 1-27e-56 Length: 2002
 Score: 623.00 Matches: 155
 Percent Similarity: 48.25% Conservative: 79
 Best Local Similarity: 31.96% Mismatches: 175
 Query Match: 20.91% Indels: 76
 DB: 1 Gaps: 18

US-09-712-338-2 (1-555) x US-08-608-267-3 (1-2002)

Qy 45 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 64

Db 533 GCCTATGATCTCGGGTCAAGAGACCGATCCCTAGCTCTCTGGCATC-----GAC 583

Qy 65 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 83

Db CCTGGCGTAAGCAGTACACCGGTTATCTCGATGACACGACGACGACGATCTGTTTC 643
 Qy 84 PheTyrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTyrLeuAsn 103
 : : : : :
 Db 644 TACTGGTTCTTCGAGTCTCGCAATGACCCCGAGAATACCCCTGTTGTTCTGTGGCTGAAC 703
 Qy 104 GlyGlyProGlySerAspSerLeuIleGlyLeuGluGluLeuGlyProCysHisVal 123
 : : : : :
 Db 704 GGTGGCCCTGGATGCTCTTCCTCACCGGTTCTTTATGAGAGCTCGGCCCTAGCAGCATC 763
 Qy 124 AsnSerThrPheAspTyrIleAsnProHisSerTyrAsnGluValSerAsnLeu 143
 : : : : :
 Db 764 AACAGAGATCCACCGCTCTACACGACTACGCTTGGAATCCACGCGTCCGTGATC 823
 Qy 144 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 163
 : : : : :
 Db 824 TTCCTTGACCGACGCTCTCAAGTCTGGTTACTCTTACAGCAACTCTGCT----- 871
 Qy 164 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 183
 : : : : :
 Db 871 ----- 871
 Qy 184 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIle 203
 : : : : :
 Db 872 -----GTGACGACACCGCTGCTGCTGGCAGGACGCTCTATGCCCTTG 913
 Qy 204 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 223
 : : : : :
 Db 914 CTTACCCCTCTCTTCAAAACAATTCCCC-----GAGTATGCCAAGCAGGACTTC 961
 Qy 224 SerLeuTyrThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyr 243
 : : : : :
 Db 962 CACATTGCCGGTGAATCTCTATGCTGTCATATATCCCGCTTTGCTCGGAGATTG 1021
 Qy 244 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 263
 : : : : :
 Db 1022 TCTCACAGAGACGCG-----AACATCAACCTGCAGTCC 1054
 Qy 264 LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPhe 283
 : : : : :
 Db 1055 GTTCTTATTGGCAACGGTCTCACCGCGTCTCACTCAGTACGAGTACTACCGTCCCATG 1114
 Qy 284 AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLys 303
 : : : : :
 Db 1115 GCCTGTGGTGACGGT---GGTTACCCAGCTGCTGGACGAG---GGCTCCTGCCAGGCC 1168
 Qy 304 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn 323
 : : : : :
 Db 1169 ATGGACACGCGCTTCTCTCGC---TGCCAGTCTATGATTGAGTCTTGTATAGTTCGGAG 1225
 Qy 324 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 343
 : : : : :
 Db 1226 AGCGCT-----TGGGTTTGTGTCGCGCTCCATCTACTGTAAACAACGCC 1270
 Qy 344 ValGluGlyProTyrTyrAlaPheAlaArgGlyValTyrAspIleArgHisProTyr 363
 : : : : :
 Db 1271 CTCCTTGCCCTTACGAGCGC---ACCGACAGAACGCTCTACGATGTTGTTAGTAGTC 1327
 Qy 364 AspAspPro-----ThrProProSerTyrTyrAsnLysPheLeuAlaLys 378
 : : : : :
 Db 1328 GAGGATAGCTCCAACTCTGCTACTCGGCCATGGGCTACGTCAGCAGCTACTGTAAACAAG 1387
 Qy 379 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsp 397
 : : : : :
 Db 1388 ACCGAGGTGATGAGGCTGTGGCGTGGAGTCAACGGCTACGCTCGGTGCAACTTTGAC 1447
 Qy 398 ValTyrTyrAlaPheGlnThrGlyAspPheValTrpPro-----AsnPheIleGlu 415
 : : : : :
 Db 1448 ATCAACCGCACTCTCTCTCCAGGTGACTGTAGTACGCTTACCAACCGCTCTCGTTCGG 1507
 Qy 416 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 434
 : : : : :
 Db 1508 GGACTCTCTGAG-----CAGATCCCTGTC-----CTGATCTACGTGGTGACGCC 1552

QY 435 AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerCln 454
 Db 1553 GATTTCATCTCAACTGGCTGGCAACAGACCTGGACTGAGCCCTTGAGTGGCCCGGA 1612
 QY 455 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 469
 Db 1613 CAGGCTGAATATGCTCCGCTTAAGCTGGAGGACCTGGTGGTGGAGATGACACAG 1672
 QY 470 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 489
 Db 1673 GGCAGAGATGCGCCAGGTCAAGTCCATGGCAACTTCACCTTCATGCGTCTATGGC 1732
 QY 490 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 509
 Db 1733 GGTGGCCACATGTCCTGGATGGACCAACCCGAGTCGAGTCTTGATTCCTCAACCCGTGG 1792
 QY 510 IlePheGly---Trp 513
 Db 1793 TTGGGAGGTGAATGG 1807

RESULT 4

US-08-608-452-3
 ; Sequence 3: Application US/08608452
 ; Patent No. 5693510
 ; GENERAL INFORMATION:
 ; APPLICANT: Yaver, Debbie Sue
 ; APPLICANT: Thompson, Sheryl Ann
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 56935100 No. 5693510disk of No. 5693510th America, Inc.
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/608,452

FILING DATE: 28-FEB-1996

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/309,341

FILING DATE: 28-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lowney, Karen A.

REGISTRATION NUMBER: 31,274

REFERENCE/DOCKET NUMBER: 4247.000-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 867 0123

TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2002 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE:

ORGANISM: Aspergillus niger

FEATURE:

NAME/KEY: intron

LOCATION: 349..411

FEATURE:

NAME/KEY: CDS

LOCATION: join (348..412)

US-08-608-452-3

Alignment Scores:

Pred. No.: 1.27e-56 Length: 2002
 Score: 623.00 Matches: 155
 Percent Similarity: 48.25% Conservative: 79
 Best Local Similarity: 31.96% Mismatches: 175
 Query Match: 20.91% Indels: 76
 DB: 1 Gaps: 18

US-09-712-338-2 (1-555) x US-08-608-452-3 (1-2002)

QY 45 AlaAsnValThrIleAlaGlyLysGluProGlyAlaGluGlyValCysGluThrThr 64
 Db 533 GCCTATGATCTCAGGTCACAGAGACCGATCTCTTGGCATC-----GAC 583
 QY 65 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 83
 Db 584 CTTGGCGTAAAGCAGTACACCGGTTATCTCGATGACACGAGACGACGACATCTGTTTC 643
 QY 84 PheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTyrLeuAsn 103
 Db 644 TACTGGTCTTCGAGTCTCGCAATGACCCGAGAAATGACCTGTGTCTGTGCTCAAC 703
 QY 104 GlyGlyProGlySerAspSerLeuIleGlyPheGluGluLeuGlyProCysHisVal 123
 Db 704 GGTGGCCCTGGATGCTCTCCCTACCGGTCCTTTTCATGAGCTCGGCCCTAGCAGCATC 763
 QY 124 AsnSerThrPheAspTyrIleAsnProHisSerTyrAsnGluValSerAsnLeuLeu 143
 Db 764 AACAGAGATCCAGCGGTCTACAACGACTACGCTTGAATCCACGCGTCCGTCATC 823
 QY 144 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 163
 Db 824 TTCCTTGACCGCTCTGCAACGTCGGTTACTCTTACAGCAACTCTGCT----- 871
 QY 164 AsnProValThrGlyValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 183
 Db 871 ----- 871
 QY 184 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaAlaAlaTzPzGluIle 203
 Db 872 -----GTCAGCAGACCGGTTGCTGCTGGCAGGACGCTGATGCTTGG 913
 QY 204 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 223
 Db 914 CTTACCTCTTCTTCAACAATTCGCC-----GAGTATGCCAAGCAGGACTTC 961
 QY 224 SerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyr 243
 Db 962 CACATTGCCGGTGAATCCTATGCTGCTACTATATCCCGTCTTTGCTTCGGAGATTGG 1021
 QY 244 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 263
 Db 1022 TCTCACAGAGAGCGC-----AACATCAACCTCGAGTC 1054
 QY 264 LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPhe 283
 Db 1055 GTCTTATTTGGCAGCGGCTCTCACCAGCGTCTCACTAGTACGAGTACTACCGTCCCATG 1114
 QY 284 AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLys 303
 Db 1115 GCCTGTGTGACGGT---GGTTACCGAGCTGCTTGGACGAG---GGCTCTCGCCAGGCC 1168
 QY 304 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerTyrCysLysGlnThrAsn 323
 Db 1169 ATGGACAACGCCCTCTCTCGC---TGCCAGCTATGATTGAGTCTGTGTATAGTTCGAG 1225
 QY 324 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 343
 Db 1226 AGCGCT-----TGGGTTTGTCTCCGGCTCCATCTACTTAACAACGCC 1270
 QY 344 ValGluGlyProTyrTyrAlaPheAlaGlyValGlyValTyrAspIleArgHisProTyr 363
 Db 1271 CTCCTTGCCCTTACCAGCGC---ACCGACAGAACGCTACGATGTTTCGTGGTAAGTGC 1327

SEQUENCE CHARACTERISTICS:
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE: Aspergillus niger
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: intron
LOCATION: 349..411
FEATURE:
NAME/KEY: CDS
LOCATION: join (348..412)
US-08-608-224-3

Alignment Scores:
Pred. No.: 1,27e-56 Length: 2002
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best local Similarity: 31.36% Mismatches: 175
Query Match: 20.91% Indels: 76
DBs: 1 Gaps: 18

US-09-712-338-2 (1-555) x US-08-608-224-3 (1-2002)

QY 45 AlaAsnValThrIleArgTyrSerProGlyAlaGluGlyValCysGluThrThr 64
DB 533 GCCTATGATCTCAGGTCAGGAAGACCGATCTCTCTGGGATC-----GAC 583
QY 65 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 83
DB 584 CTGCGGTAAAGACAGTACACCGTTATCTCGATGACACGAGACAGACATCTGTTC 643
QY 84 PheTyrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsn 103
DB 644 TACTGTTCTTCGAGTCTCGCAATACCCCGAGAATGACCTGTGTGTTCTGTGGTGAAC 703
QY 104 GlyGlyProGlySerAspSerLeuIleGlyLeuGluGluGlyProCysHisVal 123
DB 704 GGTGGCCCTGGATGCTCTTCCCTCACCGGTCTTTTCATGAGCTCGCCCTAGCAGCATC 763
QY 124 AsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 143
DB 764 AACAGAGATCCAGCGGTCTACACGACTAGCTTGGAACTCCAAACGCTCGGTATC 823
QY 144 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 163
DB 824 TTCCCTGACGAGCTGTCAACGTCGGTTACTCTTACAGCAACTCTGCT----- 871
QY 164 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 183
DB 871 ----- 871
QY 184 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGluIle 203
DB 872 -----GTCAGCAGACCGGTGCTGCTGGCAAGCAGCGTCTATGCGTTG 913
QY 204 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 223
DB 914 CTTACCTCTCTTCTCAACAAATTTCCCC-----GAGTATGCCAAGCAGGACTTC 961
QY 224 SerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyr 243
DB 962 CACATTGCCGGTGAATCTCTATGCTGCTACTATATCCCGCTCTTTCGCGAGATTG 1021
QY 244 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 263
DB 1022 TCTCAAGAGAGAGCGC-----AACATCACTACCTCGAGTCC 1054
QY 264 LeuGlyIleLeuAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTrpGluPhe 283
DB 1055 GTTCTATTGGCAACGGTCTCACCAGCGGTCTCAGTACGAGTACTACCGTCCCATG 1114

QY 364 AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 378
DB 1328 GAGTAGTCCCAACCTCTGCTACTCGGCCCTAGCTACGTCAGGACTACCTGAACAAG 1387
QY 379 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsp 397
DB 1388 ACCGAGGTCAATGAGGTGTTGGCGCTGAGTCAACGGCTAGCTCGTGAACCTTTCAC 1447
QY 398 ValTyrTyrAlaPheGlnThrGlyAspPheValTrpPro-----AsnPheIleGlu 415
DB 1448 ATCAACCCCACTCTCTTCCACGCTGACTGATGAAGCCCTACACCGTCTCGTTCG 1507
QY 416 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 434
DB 1508 GGACTCCTGGAG-----CAGATCCCTGTC-----CTGATCTACGCTGGTGAACGCC 1552
QY 435 AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGln 454
DB 1553 GATTTCACTGCAACTGCTGGCAACAGGCTGACTGAAGCCCTTGAAGTGGCCCGGA 1612
QY 455 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 469
DB 1613 CAGGCTGAATATGCTCCGCTAAGCTGGAGGACCTGCTGTCGAGATCGACACAAG 1672
QY 470 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 489
DB 1673 GCAAGAGATCGGCCAGGTCAAGTCCCATGCAACTTCACTTCATGCTGCTCTATGCG 1732
QY 490 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 509
DB 1733 GGTGGCCCATGTCCTCCGATGATGACCAACCGGATCGAGTCTGATTTCTTCAACCGCTGG 1792
QY 510 IlePheGly---Tyr 513
DB 1793 TTGGGAGGTAATGG 1807

RESULT 5

US-08-608-224-3
Sequence 3, Application US/08608224
Patent No. 5705376
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Thompson, Sheryl Ann
TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 57053760 No. 5705376disk of No. 5705376th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/608,224
FILING DATE: 28-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:

QY 284 AlaValAsnAsnThrTyrGlyLeuAlaValAsnGluThrValTyrAsnTyrMetLys 303
Db 1115 GCCTGTGGTGACGGT---GGTATCCAGCTGCTTGGACAGAG---GGCTCTGCCAGGCC 1168
QY 304 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuSerThrCysLysGlnThrAsn 323
Db 1169 ATGGACAACGCCCTCTCTCGC---TGGCAGTCTATGATTGAGTCTTCTATAGTTCGGAG 1225
QY 324 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 343
Db 1226 AGCGCT-----TGGGTTGTGTCCCGCCCTCCATCTACTCTGTAACAACGCC 1270
QY 344 ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr 363
Db 1271 CTCTCTGCCCTTACCAGCC---ACGGACAGACGCTTACGATGTCTGTGTAGTGC 1327
QY 364 AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 378
Db 1328 GAGGATAGTCCAACTCTCTACTCGGCCATGGCTACGCTACGAGCTACTCTGAACAAG 1387
QY 379 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsp 397
Db 1388 ACCGAGTCAATGAGGCTGTGGCGTGAAGTCAACGGCTACGACTCGTGCACACTTTCAC 1447
QY 398 ValTyrTyrAlaPheGlnGlnThrGlyAspPheValTyrPro-----AsnPheIleGlu 415
Db 1448 ATCAACGGCACTCTCTTCCACGGTGACTGGATGAAGCCCTACCACTCTCGTTCG 1507
QY 416 AspLeuGluLeuLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 434
Db 1508 GGACTCTCTGAG-----CAGATCCCTGTC-----CTGATCTACGCTGTGTGACGCC 1552
QY 435 AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGln 454
Db 1553 GATTTCATCTGCACTGGCTGGGCAACAGCGCTGGACTGAAGCCCTTGAGTGGCGCCGGA 1612
QY 455 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 469
Db 1613 CAGGCTGAATATGCTCCGCTAGCTGGAGGACCTGGTGTGTCAGATGAGCAAG 1672
QY 470 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 489
Db 1673 GGCAAGAAGATCGCCGAGGTCAAGTCCCATCCCACTTCACTTCATGCTCTCTATGCG 1732
QY 490 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 509
Db 1733 GTGGGCCACATGTTCCGATGACCAACCCGAGTCGAGTCTTGATTTCTCAACCGCTGG 1792
QY 510 IlePheGly---Trp 513
Db 1793 TTGGGAGGTGAATGG 1807

RESULT 6

US-08-967-149-3
; Sequence 3, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; NUMBER OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5939305 of No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,149
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,452
FILING DATE: 28-FEB-1996
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowrey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247.000-US
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: intron
LOCATION: 349..411
FEATURE:
NAME/KEY: CDS
LOCATION: join (348..412)
US-08-967-149-3

Alignment Scores:
Pred. No.: 1,278-56 Length: 2002
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 20.91% Indels: 76
DB: Gaps: 18

US-09-712-338-2 (1-555) x US-08-967-149-3 (1-2002)

QY 45 AlaAsnValThrIleArgTyrLysGluProGlyValAlaGluGlyValCysGluThrThr 64
Db 533 GCCTATGATCTCAGGTCAGAGACCGATCTAGCTCTCTGGCATC-----GAC 583
QY 65 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 83
Db 584 CTGGCGTAAAGCAGTACACCGGTTATCTCGATGACAGAGAACGACAGCATCTGTTTC 643
QY 84 PheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsn 103
Db 644 TACTGGTTCTTCAGTCTCGCAATGACCCGAGATGACCCCTGTGTTCTGTGGCTGAAC 703
QY 104 GlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisVal 123
Db 704 GGTGGCCCTGGATGCTTCTCCCTCACCCTCTTTTCATGAGCTCGGCCCTAGCAGATC 763
QY 124 AsnSerThrPheAspAspTyrIleAsnProHisSerTyrAsnGluValSerAsnLeuLeu 143
Db 764 AACAAAGAGATCCAGCCGGTCTTACAAACGACTACGCTTGAACCTCCAAACGCTCGGTGATC 823
QY 144 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 163
Db 824 TTCTTGACCGCTGTCAAGCTCGGTACTCTTACAGCAACTCTGCT-----871
QY 164 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 183
Db 871 -----871

QY	103	AsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHis	122
DB	687	AACGGTGGCCCTGGGTCTCTCCCTCACCGGTCTCTCATGGAGCTTGGCCCTAGCAGC	746
QY	123	ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu	142
DB	747	ATCAACAAGAAAGATCCAGCCGGTCTACAATGACTACGCTTGGAATCCCAACGCGTCGCTG	806
QY	143	LeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySer	162
DB	807	ATCTCTCTTGACCAACCCCTGTAATGTCGGTACTCTCTACAGTAAGTCTGCT	857
QY	163	IleAsnProValThrGlyValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr	182
DB	857	-----	857
QY	183	ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGlu	202
DB	858	---GTCAGCAGACAGGTCGCTGCTGGCAAGAGCGTCTATGCC	896
QY	203	IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerIlyAsp	222
DB	897	TTCGTTACCCCTCTCTTCAACAATGCC-----GAGTATGCTTAAGCAGGAC	944
QY	223	PheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPhe	242
DB	945	TTCCACATGCGCGTGAATCTTATCTGCTCACTATATCCCGCTCTTCGCTTCGGAGATC	1004
QY	243	TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn	262
DB	1005	CTGTCTCACAGAAGCGC-----AACATCAACCTCGAC	1037
QY	263	SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGlu	282
DB	1038	TCCGTTCTCATTTGGCAACGGTCTCACCAGCGATACCCAGTACGAGTACTACCGTCCC	1097
QY	283	PheAlaValAsnAsnThrTyrGlyIleIleAlaValAsnGluThrValTyrAsnTyrMet	302
DB	1098	ATGSCCTGCGGTGACGGCGGTATCCACGCTCTTGGACAGAGACTCTGCCAGTCCATG	1157
QY	303	LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr	322
DB	1158	GAC-----AACGCTCTCTCCTCGC-----TGCCAGTCTATGATGAGTCTTGCTACAGTTCC	1208
QY	323	AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp	342
DB	1209	GAGAGCGCT-----TGGGTTTGTTCCCGCCCTCCATCTACTGTAAACAAC	1253
QY	343	AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro	362
DB	1254	GCCTCTCTGCCCTTACCGGC-----ACTGGCAGACGCTCTATGATGTCCTGGTGAAG	1310
QY	363	TyrAspAspPro-----ThrProSerSerTyrTyrAsnLysPheLeuAla	377
DB	1311	TGGAGGAGTAGCTCTAACCTTTGCTACTCGGCTATGGGCTACGTCAGCGACTACCTGAAC	1370
QY	378	LysAspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsn	396
DB	1371	AAGCCCGAGTCAATCAGGCTGTGTGGCGTGAAGTCAACGGCTACGACTCGTGCACTTT	1430
QY	397	AspValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIle	414
DB	1431	GACATCAACCGCAACTTCTCTTCCAGGTGACTGGATGAAGCCCTACCAACCGCTCGTT	1490
QY	415	GluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAsp	433
DB	1491	CGGGGACTCCTGGAG-----CAGATCCCTGTC-----TTGATCTATCGCGTGAT	1535
QY	434	AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSer	453
DB	1536	GCTGATTTCAITTGCAACTGGGTGGGCAACAAGGCTTGACCTGAAGCCCTGGAGTGGCC	1595

```

QY 454 GlnAlaalaGlnPheArgSerAlaGlyTyrThrProLeuLysVal----- 458
Db 1596 GGACAGGCTGAATATGCGCTCCGCTAGCTGGAGGATCTGCTCATTTGCGACAATGAGCAC 1655
QY 469 AsnGlyValGluTyrGlyGluThrArgGluTyr-GlyAsnPheSerPheThrArgValTyr 488
Db 1656 ACGGGCAAGAGATTGGCCAGGTTAAGTCCATGGCAACTTCACCTTCATGCGCTCTCTAT 1715
QY 489 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 508
Db 1716 GGTGGTGGCCACATGTCGCCGATGCGACCACCCGAGTCGAGTTCGAGTTCTTCAACCGC 1775
QY 509 ThrIlePheGly---Tyr 513
Db 1776 TGGTTGGGAGGTGAATGG 1793

RESULT 8
US-08-608-267-1
: Sequence 1, Application US/08608267
: Patent No. 5688663
: GENERAL INFORMATION:
: APPLICANT: Yaver, Debbie Sue
: APPLICANT: Thompson, Sheryl Ann
: TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.
: STREET: 405 Lexington Avenue, Suite 6400
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/608,267
: FILING DATE: 28-FEB-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/309,341
: FILING DATE: 20-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Lowney, Karen A.
: REGISTRATION NUMBER: 31,274
: REFERENCE/DOCKET NUMBER: 4247.000-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212 867 0123
: TELEFAX: 212 867 0298
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2068 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: ORIGINAL SOURCE:
: ORGANISM: Aspergillus niger
: FEATURE:
: NAME/KEY: intron
: LOCATION: 572..632
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join (571..633)
: IS-08-608-267-1

```

Alignment Scores:	
Pred. No.:	1.69e-53
Score:	594.00
Percent Similarity:	46.44%
Best Local Similarity:	30.63%
Length:	2068
Matches:	155
Conservative:	80
Mismatches:	174

Query Match:	19.94%	Indels:	98
DB:	1	Gaps:	18
US-09-712-338-2 (1-555) x US-08-608-267-1 (1-2068)			
QY	45	AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr	64
DB	457	GCCTATGATCTCAGGGTCAAGAAGACCGATCTGGCTCTCTTGGCATC	507
QY	65	ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu	79
DB	508	CCCGCGTGAACAGTACACCGGTATCTCGATGACACGA-GAATGATACCATTTGTT	566
QY	80	-----SerHisThrPhePhe-----	84
DB	567	CTACGTAAACACACCTTGTTCAAGATCAGCTTTTATATGCTCTGGATATCTAACGCA	626
QY	85	-----TrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeu	102
DB	627	ACTTAGTGTTCTTCAGTCTCGCAATACCCGAGATGATCCGTTGTTCTGTGGCTG	686
QY	103	AsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluGlyProCysHis	122
DB	687	AACGGTGGCCCTGGGCTCTTCCCTCACCGGTCTCTTATGAGCTTGGCCCTAGCAGC	746
QY	123	ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu	142
DB	747	ATCAACAAGAAGATCCAGCGGTCTACATGACTACGCTTGAACCTCCAACGCTCCGTG	806
QY	143	LeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySer	162
DB	807	ATCTTCTTGAACGAGCTCTCAATGCTGCTTCTCTACAGTAACTCTCT	857
QY	163	IleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr	182
DB	857	-----	857
QY	183	ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaLapTrpGlu	202
DB	858	-----GTACGACACACGCTGCTGCTGGCAAGACGCTATGCC	896
QY	203	IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp	222
DB	897	TTGCTTACCTCTCTTCAACAATTCCTCC	944
QY	223	PheSerLeuThrProGlySerTyrGlyHisTyrGlyProAlaPheAsnHisPhe	242
DB	945	TTCACATTCGCGGTGAATCTTATGCTGCTACATATATCCCTCTTCGTTCCGAGATC	1004
QY	243	TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn	262
DB	1005	CTGTCTCACAGAAGCGC-----AACATCAACCTGCAG	1037
QY	263	SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGlu	282
DB	1038	TCCGTTCTCATTCGCAACGCTCTCACCGCGATACACCAGTACGAGTACTACCGTCCC	1097
QY	283	PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet	302
DB	1098	ATGGCTGCGGTGACGCGGTATCCAGCTGTCTTGACGAGAGCTCTCCGATCCATG	1157
QY	303	LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr	322
DB	1158	GAC-----AACGCTCTTCTCGC-----TGCGAGTCTATGATTCAGTCTTCACAGTTC	1208
QY	323	AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp	342
DB	1209	GAGAGCGCT-----TGCGTTTGTCCCGGCTCCATCTACTGTAACAAC	1253
QY	343	AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro	362
DB	1254	GCCCTCTTGGCCCTTACACGCG-----ACTGGGACAGACGCTATGATGTCGCGTGAAG	1310

RESULT 9

US-08-608-452-1
 ; Sequence 1, Application US/08608452
 ; Patent No. 5693510
 ; GENERAL INFORMATION:
 ; APPLICANT: Yaver, Debbie Sue
 ; APPLICANT: Thompson, Sheryl Ann
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 56935100 No. 5693510disk of No. 5693510th America, Inc.
 ; STREET: 405 Lexington Avenue, Suite 6400
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patcin Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/608,452
 ; FILING DATE: 28-FEB-1996
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/309,341
 ; FILING DATE: 20-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lowney, Karen A.
 ; REGISTRATION NUMBER: 31,274
 ; REFERENCE/DOCKET NUMBER: 4247,000-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 867 0123
 ; TELEFAX: 212 867 0298
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 435
FILING DATE: 28-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2068 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: intron
LOCATION: 572..632
FEATURE:
NAME/KEY: CDS
LOCATION: Join (571..633)
US-08-608-224-1

Alignment Scores:
Pred. No.: 1.69e-53
Score: 594.00
Percent Similarity: 46.44%
Best Local Similarity: 30.63%
Query Match: 19.94%
DB: 1
Gaps: 18
Indels: 98
Matches: 155
Length: 2068

US-09-712-338-2 (1-555) x US-08-608-224-1 (1-2068)

QY	45	AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr	64
DB	457	GCTATGATCTAGGTCAGAGACCGATCTCGTCTCTTGGCATC-----GAC	507
QY	65	ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu-----	79
DB	508	CCGGGGTGAAGCAGTACACCGGTATCTCGATGACACGA-GAATGATAAGCATTTGTT	566
QY	80	-----SerHisThrPhePhe-----	84
DB	567	CTACGTAAGCACACCTTGGTCAAGATCACGCTTTTATATGCTCTGGATATCAACGCA	626
QY	85	-----TriPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeu	102
DB	627	ACTTAGTGTTCTTCGAGTCTCGAGTACGACCCGAGAAATGATCCCGTTGTTCTGTGGCTG	686
QY	103	AsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHis	122
DB	687	AACGGTGGCCCTGGGTGCTCTCCCTCACCGGTCTCTTCATGAGGCTTGCCCTAGCAGC	746
QY	123	ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu	142
DB	747	ATCAACAAGAGATCCAGCGGTCTACAAATGACTACCTTGGAACTCCCAACGGTCCGCTG	806
QY	143	LeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySer	162
DB	807	ATCTTCTTGACCAAGCTGTCAATGTGGTTACTCTACTCTACAGTAACACTCTGCT-----	857

QY	163	IleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr	182
DB	857	-----	857
QY	183	ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGlu	202
DB	858	-----GTGAGCGACACGGTCGCTGTGGCAAGGAGCTGTATGCC	896
QY	203	IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp	222
DB	897	TTGCTTACCTCTTCTTCAAAACAATTCCTC-----GAGTATGCTAAGCAGGAC	944
QY	223	PheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPhe	242
DB	945	TTCCACATTCGGGTGAATCTTATGCTGCTCACTATATCCCGCTCTTCTGCTGGAGATC	1004
QY	243	TyrGluGlnAsnGluAlaGlnIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn	262
DB	1005	CTGCTCACAGAGAGCG-----AACATCAACCTGCAG	1037
QY	263	SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGlu	282
DB	1038	TCCGTCTCATTCGACACGGTCTCACCGACGATACACCCAGTACGAGTACTACCTGCC	1097
QY	283	PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet	302
DB	1098	ATGGCCTCGGTGACGGCGTTACCCAGCTGCTTGGAGGAGAGTCTCCGACGCTCCATG	1157
QY	303	LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr	322
DB	1158	GAC-----AACGCTCTTCTCTGC---TGCCAGTCTATGATGAGTCTTGTACAGTTCC	1208
QY	323	AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp	342
DB	1209	GAGAGCGCT-----TGGGTTTGTCTCCCGCCTCCATCTACTGTAAACAAC	1253
QY	343	AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro	362
DB	1254	GCCCTCTTGGTCCCTTACCAGCGC---ACTGGGAGAACGCTCTATGATGCTCGTGGTAAG	1310
QY	363	TyrAspAspPro-----ThrProProSerTyrTyrAsnLysPheLeuAla	377
DB	1311	TGCGAGATAGCTTAACCTTTGCTACTCGGCTATGGGTACGTACGAGCTACTCTGAAC	1370
QY	378	LysAspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsn	396
DB	1371	AAGCCGGAAGTCAATCGAGGCTGTGGCGCTGAGTCAACGGCTACGACTCGTGCAACTTT	1430
QY	397	AspValTyrTyrAlaPheGlnThrGlyAspPheValTrpPro-----AsnPheIle	414
DB	1431	GACATCAACCGCAACTTCTCTTCCAGGTGACTGGATGAAGCCCTACCAACCGCTCGTT	1490
QY	415	GluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAsp	433
DB	1491	CCGGGACTCTCTGAG-----CAGATCCCTGTC-----TTGATCTATGCGGTGAT	1535
QY	434	AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSer	453
DB	1536	GCTGATTTCTATTTGCAACTTGGTGGCAACAGGCGCTGGACTGAAGCCCTGGAGTGCC	1595
QY	454	GlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----	468
DB	1596	GGACAGCTGAATATGCTCCGCTGAGCTGGAGGATCTGGTCATTGTCGACATAGACAC	1655
QY	469	AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr	488
DB	1656	ACGGGCAAGAAGATTGCCAGGTAAAGTCCCATGGCAACTTCACTTCACTGCGTCTCTAT	1715
QY	489	GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg	508
DB	1716	GGTGGTGGCCACATGTTCCCGATGGACCCGAGTCGAGTCTCGAGTCTTCAACCGC	1775
QY	509	ThrIlePheGly---Trp	513


```

Db      1776 TGGTTGGAGGTGAATGG 1793
      ... ||| |||
RESULT 11
US-08-967-149-1
; Sequence 1, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59393050 No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join (571..633)
US-08-967-149-1
Alignment Scores:
Pred. No.: 1-69e-53 Length: 2068
Score: 594.00 Matches: 155
Percent Similarity: 46.44% Conservative: 80
Best Local Similarity: 30.63% Mismatches: 174
Query Match: 19.94% Indels: 98
DB: 2 Gaps: 18
US-09-712-338-2 (1-555) x US-08-967-149-1 (1-2068)
QY      45 AlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 64
      |||
Db      457 GCCTATGATCTCAGGTCAGAGACCGGATCTGGCTCTCTGGGATC-----GAC 507
QY      65 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu----- 79

```

Db 1431 GACATCACCGCAACTTCTCTCCACGGTGACGTGATGAAGCCCTACCAACGGCTCGTT 1490
QY 415 GluAspLeuGluGluLeuAlaLeuProValArgValSerLeuIleTyr---GlyAsp 433
Db 1491 CCGGACCTCTGGAG-----CAGATCCCTGTC-----TTGATCTATCCCGTGAT 1535
QY 434 AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSer 453
Db 1536 GCTGATTTCATTGCAACTGGCTGGGCAACAGGCTGAGACTGAAGCCCTGGAGTGGCC 1595
QY 454 GlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal----- 468
Db 1596 GGACAGGCTGATATGCTCCCTCCCTGAGTGGAGGATCTGGTCAATTGTGCACATGAGCAC 1655
QY 469 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 488
Db 1656 ACGGCAAGAGATGGCCAGGTAAAGTCCCATGGCACTTCACTTCATGCTCTCTAT 1715
QY 489 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 508
Db 1716 GGTGGTGGCCACATGCTCCGATGGACACCGCCGAGTCGAGTCTCGAGTCTTCAACCGC 1775
QY 509 ThrIlePheGly---Tyr 513
Db 1776 TGGTTGGGAGTGAATGG 1793

RESULT 12
US-09-640-305-3
; Sequence 3, Application US/09640305
; Patent No. RE37447
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; Fournier, Alain
; Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/640,305
; FILING DATE: 16-Aug-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673
; FILING DATE: 06-FEB-1995
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

ORIGINAL SOURCE:
; ORGANISM: Kluyveromyces lactis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 387..1862
; OTHER INFORMATION: /product= "K. lactis protease C
; gene"
; /gene= "K1.PRC1"
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-640-305-3
Alignment Scores:
Pred. No.: 1,21e-50 Length: 2503
Score: 568.50 Matches: 151
Percent Similarity: 44.65% Conservative: 66
Best local Similarity: 31.07% Mismatches: 184
Query Match: 19.08% Indels: 85
DB: 1 Gaps: 14
US-09-712-338-2 (1-555) x US-09-640-305-3 (1-2503)
QY 45 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 64
Db 570 GCTTATTCATTGAGAAATTAAACCTTTGGATCCCAATCTCTTGGGCTT----- 617
QY 65 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePhe 84
Db 618 GATACCGTGAACAATGGTCGGGATATTAGATTACACGAGTCAAAACACTCTTTTAT 677
QY 85 TrpPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGly 104
Db 678 TGGTTTTTGAGCTAGAAATGACCCAGAGATGACCCAGTACTATGTTAAACGCT 737
QY 105 GlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsn 124
Db 738 GGTCTCGCTGTTCT 797
QY 125 SerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuPhe 144
Db 798 GCTGATTTGAAACCCATTTATACCCCTACTCTTGGAATTCACACCTCTCTGTATTC 857
QY 145 LeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsn 164
Db 858 CTAGATCAGCCTGT 896
QY 165 ProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThr 184
Db 896 ----- 896
QY 185 IleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGluIleLeu 204
Db 897 -----TCTAAAGTGTCTACTCAGATGACGCTGCCAAGACGTTTACATATTCTTA 947
QY 205 GlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSer 224
Db 948 GATTTCCTCTTGAAGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 995
QY 225 LeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGlu 244
Db 996 ATCTCCGGTGAATCATAACGCGGTCATTATTATCCCAAGATTCTCTCTCTCTCTCT 1043
QY 245 GlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeu 264
Db 1044 -----CAGATTGCTGTAGTGCATGCTGAGATTCCTCTCTCTCTCTCTCTCT 1094
QY 265 GlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAla 284
Db 1095 TTAATTGGAATGGATTTACTGACCCACTGACTCAATACCAATATTACGAGCCGATGCC 1154
QY 285 ValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPhe 304
Db 1155 TGTGTGAAGGT---GGTTATCCAGCGGTGTTGAA----- 1187

```

QY 305 AlaAsnGlnMetProAsnGlyCysGlnAsp----- 314
Db 1188 -----CCGGAAGATTCTTAGATATGAATAGAACTACCTCTATGCTATCG 1235
QY 315 LeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAla 334
Db 1236 CTTGTGGACCTGTTACAAGTCCCAT-----TCTGTTTCTCTGTGTG 1280
QY 335 GluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArg 354
Db 1281 TTGGCTGACCTTATTGTGAACAACAGATTACTGG---GTTATGAAATCAGGTAGG 1337
QY 355 GlyValTyrAspIleArg-----HisProTyrAspAspPro-----Thr 367
Db 1338 AACCTTACGATATTAGATCTAAGTGAACAGAGGATGATCCGGTGCCTGTATCAG 1397
QY 368 ProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyVal 387
Db 1398 GAAGAAATTTATCTCTGATTCTTGAATCAGGAGGAAGTTCAAAGAGCTTTAGGGACT 1457
QY 388 AsnIleAsnTyrThrGlnSer---AsnAsnAspValTyrTyrAlaPheGlnGlnThrGly 406
Db 1458 GATGTGAGTTCTTTCCAAAGTTGTAGCTCGATCGGTATCGGTTTCGGATTCACTGGC 1517
QY 407 AspPheValTrpProAsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArg 426
Db 1518 GATGGACCGAGCCCA---TTCCACCACTAGCTCGCAGAACTTCTTGATCAAGATATCAAT 1574
QY 427 ValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlnAlaVal 446
Db 1575 GTCTTGATATATGACGCGGATAAGGATTATTTGTAATTTGGCTAGGAAATCTCGCTGG 1634
QY 447 SerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeu 466
Db 1635 ACTGAAATAATTTGAATCGAGGTATACGAGAGGTATATAAATAACAAGTTTGAGAACTGG 1694
QY 467 LysValAsnGlyValGlu-----TyrGlyGluThrArgGluTyrGlyAsnPheSerPhe 484
Db 1695 AAGAGTGAAGAAACAGATGAGACCATTTGGCGAAACCAATCTTATGGCCGCTAACTTAC 1754
QY 485 ThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGln 504
Db 1755 TTGAGATCTATGATGCTGACACATGTTCTCTCAGCACCACTGAAATTCATTACAA 1814
QY 505 LeuPheAsnArgThrIle 510
Db 1815 ATGGTGAATTCATGGATT 1832

RESULT 13
US-08-360-673-3
; Sequence 3, Application US/08360673
; Patent No. 5679544
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3043
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673

```

```

; FILLING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Kluyveromyces lactis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 387..1862
; OTHER INFORMATION: /product= "K. lactis protease C
; OTHER INFORMATION: gene"
; OTHER INFORMATION: /gene= "K1.PRC1"
; US-08-360-673-3

```

```

Alignment Scores:
Pred. No.: 1,21e-50 Length: 2503
Score: 568.50 Matches: 151
Percent Similarity: 44.65% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 184
Query Match: 19.08% Indels: 85
DB: 1 Gaps: 14

```

US-09-712-338-2 (1-555) x US-08-360-673-3 (1-2503)

```

QY 45 AlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 64
Db 570 GCTTATTCATTGAGAAATTAACCTTTGGATCCAAATCTCTTGGCGTT----- 617
QY 65 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePhe 84
Db 618 GATACCGTGAACAATGTCGGGATATTAGATTACCAGGACTCAAAACACTTCTTTTAT 677
QY 85 TrpPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGly 104
Db 678 TGGTTTTGTAGCTAGAAATGACCCAGAGAATGACCCAGTACTATGTTAAACGGT 737
QY 105 GlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsn 124
Db 738 GGTCTGGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 797
QY 125 SerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPhe 144
Db 798 GCTGATTTGAACACCATTTATACCCCTACTCTTGGAAATCCACGCTTCTGTGATATC 857
QY 145 LeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsn 164
Db 858 CTAGATCATGCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 896
QY 165 ProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThr 184
Db 896 -----
QY 185 IleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeu 204
Db 897 -----TCTAAAGTGTCTACTACAGATGACGCTGCCAAAGAGCTTTTACATATCTTA 947

```

GENERAL INFORMATION:
 APPLICANT: Bredam, Klaus
 APPLICANT: Keilland-Brandt, Morten
 APPLICANT: Mortensen, Uffe
 APPLICANT: Olesen, Kjeld
 APPLICANT: Stennicke, Henning
 APPLICANT: Wagner, Fred
 TITLE OF INVENTION: CUSTOMIZED PROTEASES
 NUMBER OF SEQUENCES: 33
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchaut, Gould, Smith, Edell, Welter & Schmidt
 STREET: 3100 No. 5945329west Center, 90 S. 7th Street
 CITY: Minneapolis
 STATE: MN
 COUNTRY: U.S.A.
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/899,324
 FILING DATE: 23-JUL-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/329,892
 FILING DATE: 27-OCT-1994
 APPLICATION NUMBER: 08/144,704
 FILING DATE: 28-OCT-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Kettleberger, Denise M
 REGISTRATION NUMBER: 33,924
 REFERENCE/DOCKET NUMBER: 8648.44USCI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612/332-5300
 TELEFAX: 612/332-9081
 TELEX:
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2632 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE:
 ORIGINAL SOURCE:
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 696...2291
 OTHER INFORMATION:
 US-08-899-324-32

205 GlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSer 224
 Db GATTTGTTCTTGAAGATTCCTCATTTG-----AGAAATACGATTTCCAT 995
 225 LeuTrpThrGluSerGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGlu 244
 Db ATCTCCGGTGAATCATACCGCGGTCAATATTATACCAAGATTGCTCAT----- 1043
 245 GlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeu 264
 Db -----GAGATTCCTAGTGCATGCTAGGATTCCTCCTCAATCAATATCGFCAGTA 1094
 265 GlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAla 284
 Db TTAATGGAATGATTTACTACCCACTGCTCAATACCAATATACGAGCGGATGCC 1154
 285 ValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPhe 304
 Db TGTGCTGAAGGT---GGTTATCCAGCGGTGTGGAA----- 1187
 305 AlaAsnGlnMetProAsnGlyCysGlnAsp----- 314
 1188 -----CCGGAAGATTCTAGATATGAATGAGGAATCTACCTCTATGCCTATGC 1235
 315 LeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAla 334
 Db CTGTGGACCGCTGTACAGTCCCAT-----TCTGTTTCTCTGTGTG 1280
 335 GluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArg 354
 Db TTGGCTGACCGTTATTGTGAACAACAGATTACTGGG---GTTTATGAGAAATCAGGTAGG 1337
 355 GlyValTyrAspIleArg-----HisProTyrAspPro-----Thr 367
 1338 AACCTTACGATATAGATTAACTAGTGTGAGCGAGAGATGATTCGGTGCCTGTATCAG 1397
 368 ProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyVal 387
 Db GAAGAAATTTATCTCTGATTACTTGAATCAGGAGGAAGTTCAAAGAGCTTTAGGACT 1457
 388 AsnIleAsnTyrThrGlnSer---AsnAsnAspValTyrTyrAlaPheGlnGlnThrGly 406
 Db GATGTCAGTCTTTCGAAGGTTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 1517
 407 AspPheValTrpProAsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArg 426
 Db GTGACCGAGCCCA---TTCCACAGTACGTCGAGACTTCTTGTATCAAGATATCAAT 1574
 427 ValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaVal 446
 Db GTCTTGATATATGACGCGGATAAGGATTATATTGTAATGCTAGGAAATCTCGTTGG 1634
 447 SerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeu 466
 Db ----- 1635
 1635 ACTGAAAAATGGAATGAGGATATAACCAAGAGATATAAAAAACAAGTTTGTGAGACTGG 1694
 467 LysValAsnGlyValGlu-----TyrGlyGluThrArgGluTyrGlyAsnPheSerPhe 484
 Db AAGAGTGAAGAAACAGATGAGACCATGGCGAACCATAATCTATGCGCCGCTAATTAC 1754
 485 ThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGln 504
 Db TTGAGAAATCATGATGCTGGACATGTTCTCTCAGCAGCAACCTGAAATTCATTACAA 1814
 505 LeuPheAsnArgThrIle 510
 Db ATGGTGAATTCATGCAAT 1832

US-09-712-338-2 (1-555) x US-08-899-324-32 (1-2632)
 QY 33 ProLysAsnProThrGlyValLysThr-----LeuThrThrAlaAsnAsn 47
 Db 939 CCAAAATTCCTGAAGCAATCAAAACGAGAAAGACTGGGACTTTGTGTCAAGAAATGAC 998
 QY 48 ValThrIleArgTyr-----LysGluProGlyAlaGluGlyVal 60
 Db 999 GCAATTGAAACCAATATCAGCTTCGTGTCAACAGATTAAAGCACCTTAAATCTCGGCATT 1058

RESULT 14
 US-08-899-324-32
 ; Sequence 32, Application US/08899324
 ; Patent No. 5945329

QY 61 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu--- 79
Db 1059 -----GACCCAAATGTCACACAGTACACGGTTACTTGGATGTGGAGACGAGAC 1109
QY 80 SerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThr 99
Db 1110 AAGCATTTCTCTTTGGACTTTGAAAGTAGAAGACGATCTGCAAAAGGATCCGGTCATC 1169
QY 100 LeuTyrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGly 119
Db 1170 CTTTGGTTGACCGGGGTCCAGGTTGTCTTCACTAACCGGGCTGTCTTTGAAATTAGGA 1229
QY 120 ProCysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluVal 139
Db 1230 CCCTCATCCATTGCACCTGATTGAAACCCATCGGAAACCTTACTCTTGGACAGCAAT 1289
QY 140 SerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrVal 159
Db 1290 GCACCGGTGATCTCTCTTGACAGCCTGTCAACGTTGGTTCGTATTC----- 1340
QY 160 AspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGln 179
Db 1341 -----GGGTCTCTCA 1349
QY 180 GlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAla 199
Db 1350 GGT-----GTTTCAACACTGTGCGCGTGTGAAGAT 1382
QY 200 AlaTyrGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGln 219
Db 1383 GTCTATAACTTCTTGAGTTGTTCTTCGATCAGTTCCTCGAATACGTCACAAAGGGCCAA 1442
QY 220 SerLysAspPheSerLeuTyrThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePhe 239
Db 1443 -----GATTTCCACATCGCTGGGATCTCTAGCGCGCATTACATCCCTGTTTGGC 1496
QY 240 AsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeu 259
Db 1497 TCTGAAATTTGTCTCAAGGACAGA----- 1523
QY 260 AsnPheAsnSerLeuGlyIlelle-----AsnGlyIlelleAspGluAlaIleGlnAla 277
Db 1524 AACTTCAACTTAACCTCCGCTCTTGTATCGGAATGGCTCACTGACCCATGACTCAGTAT 1583
QY 278 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThr 297
Db 1584 AACTATTACGAACCAATGGCTGTGTGAAGT---GGCGAACCATCTGTT----- 1631
QY 298 ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 317
Db 1632 -----TTGCCCTCGGAGGAATGCTGCTATGATGGAAGACTCTTTGGAA 1673
QY 318 ThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu-----CysAlaGlu 335
Db 1674 CGTTGTTTGGGCTTGATCGAGTGTGCTGTATGACTCGCAATCGGTCTGTGTGTCTCA 1733
QY 336 AlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGly 355
Db 1734 GCTACCATTTATGTAAATACGCCCAATGGCTCTTACCAAGT---ACCGCAGAAAC 1790
QY 356 ValTyrAspIleArgHisProTyrAspAsp-----ProThrProProSer 370
Db 1791 GTTTACGATATCAGGAAGGATTGTGAAGTGCAATTTGTGCTACCAACGTTACAGAT 1850
QY 371 TyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsn 390
Db 1851 ---ATCGACGACTACTTAACCCAGGACTACGTCAAAAGAGCTGTCGGTCCGAGGTTGAC 1907
QY 391 TyrThrGlnPheSer---AsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAspPheVal 409
Db 1908 CACTAGCAATCTGTGTAATCTGATATCAACAGAAATTTCTGTTGCGGGTGTATGGATG 1967
QY 410 TrpProAsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeu 429

Db 1968 AAGCCT---TACCACACCGCCGTAACAGATCTTTGAATCAAGACCTTACCATTCTCGTA 2024
QY 430 IleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGlyGlyGlnAlaValSerLeuAla 449
Db 2025 TATGACGGCGATAAAGATTTCATCTGTAACTGTGTTGGTAAATAGGCGTGCAGCATGTC 2084
QY 450 AlaAsnTyrSerGlnAlaAlaGlnPheArgSer-----AlaGlyTyrThrProLeu 466
Db 2085 TTCCCATGGAAGTACGACGAGAAATTTGCAAGCCAAAGTACGTAACTGACTGCTTCT 2144
QY 467 LysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArg 486
Db 2145 ATCACCAGCAGAGTCCGT---GGTGAAGTCAAAATCTTACAGCACTTCACTATTGAGA 2201
QY 487 ValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPhe 506
Db 2202 GTCTTCAATGGTGGCCACATGGTTCCATTGACGTCCCTGAAACAGCCTTAAAGTATGGTT 2261
QY 507 AsnArgThrIlePheGly 512
Db 2262 AACGAATGGATCCACGGT 2279

RESULT 15

US-08-329-892B-32
Sequence 32, Application US/08329892B
Patent No. 6187579
GENERAL INFORMATION:
APPLICANT: Breddam, Klaus
APPLICANT: Keiland-Brandt, Morten
APPLICANT: Mortensen, Uffe
APPLICANT: Olesen, Kjeld
APPLICANT: Stennicke, Henning
APPLICANT: Wagner, Fred
TITLE OF INVENTION: CUSTOMIZED PROTEASE
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 3100 No. 6187579west Center, 90 S. 7th Street
CITY: Minneapolis
STATE: MN
COUNTRY: U.S.A.
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/329,892B
FILING DATE: 27-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/144,704
FILING DATE: 28-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kettleberger, Denise M
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 8648.44US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612/332-5300
TELEFAX: 612/332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 2632 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:

ORIGINAL SOURCE:

FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 696..2291
 OTHER INFORMATION:
 US-08-329-892B-32

Alignment Scores:

Pred. No.: 8,47e-46 Length: 2632
 Score: 523.50 Matches: 147
 Percent Similarity: 43.68% Conservative: 74
 Best local Similarity: 29.05% Mismatches: 200
 Query Match: 17.57% Indels: 85
 DB: 3 Gaps: 19

US-09-712-338-2 (1-555) x US-08-329-892B-32 (1-2632)

```

QY 33 ProLysAsnProThrGlyVallyThr-----LeuThrThralaAsnAsn 47
DB 939 CCAAAATTCCCTGAAGCAATCAAAAACGAAGAAGACTGGGACTTTGTGTCAAGATGAC 998
QY 48 ValThrIleArgTyr-----LysGluProGlyAlaGluGlyVal 60
DB 999 GCAATTGAAACTATCAGCTTCGTGTCAACAAGATTAGGACCCCTAAATCCTGGGCATT 1058
QY 61 CysGluThrThrProGlyVallySerTyrSerGlyTyrValAspThrSerProGlu--- 79
DB 1059 -----GACCCAAATGTCACACAGTACACGGGTACTTGGATGTGGAACGAGGAC 1109
QY 80 SerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThr 99
DB 1110 AAGCATTTCTCTTTGGACTTTTGAAGTAGAACGATCCTCGCAAGGATCCGGTCATC 1169
QY 100 LeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluGly 119
DB 1170 CTTTGGTTGAACGGGGTCCAGGTGTTCTTCACTACCGGGCTGTTCTTGAATAGGA 1229
QY 120 ProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluVal 139
DB 1230 CCCTCATCTTGGACCTGATTGAAACCCATCCGGAACCCCTACTCTTGGAAACAGCAAT 1289
QY 140 SerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrVal 159
DB 1290 GCCACCGTCATCTCTTCACACGACCTGTCAACGTTGGTTCGCTATTCCTCC----- 1340
QY 160 AspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGln 179
DB 1341 -----GGTCTCTCA 1349
QY 180 GlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAla 199
DB 1350 GGT-----GTTTCCAACACTGTCGCCGCTGGTAAGGAT 1382
QY 200 AlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGln 219
DB 1383 GTCTATAACTTCTTGAGTGTGTTCTTCGATCAGTTCCTCGAATACGTCACAAAGGCCAA 1442
QY 220 SerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePhe 239
DB 1443 -----GATTTCCACATCGCTGGGAATCCTACCGCGGCATTACATCCCTGTTTGGC 1496
QY 240 AsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeu 259
DB 1497 TCTGAAATTTGTCTCACAGGACAGA----- 1523
QY 260 AsnPheAsnSerLeuGlyIleIle-----AsnGlyIleIleAspGluAlaIleGlnAla 277
DB 1524 AACTTAACTTAACTCCGTCCTTGATCGGAAATGGCCCTACTGACCCATTGACTCAGTAT 1583
QY 278 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThr 297
DB 1584 AACTATTACGAACCAATGGCTGTGGTGAAGGT---GGCGAAACCATCTGTT----- 1631

```

```

QY 298 ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 317
DB 1632 -----TTGCCCTCGAGGAATGCTCTGATATGGAAGACTCTTTGAA 1673
QY 318 ThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu-----CysAlaGlu 335
DB 1674 CGTTGTTGGCGTTCATCGAGTCGTCGTATGACTCGCAATCGGTGCTGCTGTTGCCA 1733
QY 336 AlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGly 355
DB 1734 GCTACCAATTTATTGATAACGCCCAATTTGGCTCCTTACCAACGT---ACCGCGAGAAAC 1790
QY 356 ValTyrAspIleArgHisProTyrAspAsp-----ProThrProProSer 370
DB 1791 GTTTACGATATCAGGAAGATTGTGAAGTGGCAATTTGTGCTACCCACGTTTACAAGAT 1850
QY 371 TyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsn 390
DB 1851 ---ATCGACGACTACTTAAACACGAGCTACGTCAAAGAAAGCTGTCGGTGGGAGTTGAC 1907
QY 391 TyrThrGlnSer---AsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAspPheVal 409
DB 1908 CACTACGAATCCTGTAACCTTCGATATCAACAGAAATTTCTGTTTTCGCGGTGATTGGATG 1967
QY 410 TrpProAsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeu 429
DB 1968 AAGCCT---TACCACACCGCGCTAAACAGATCTTTTGAATCAAGACCTACCCATTCTGGTA 2024
QY 430 IleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAla 449
DB 2025 TATGCAGCGGATAAAGATTTCATCTGTTGTTGTTAAAGCGTGGACGGATGTC 2084
QY 450 AlaAsnTyrSerGlnAlaAlaGlnPheArgSer-----AlaGlyTyrThrProLeu 466
DB 2085 TTGCCATGGAGTACGACGAGAGAAATTTGCAGACCAAAAGTACGTAACCTGCTTCT 2144
QY 467 LysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArg 486
DB 2145 ATCACCAGCGAGGTGCT---GGTGAAGTCAAAATCTCTACAGCACTTCACCTATTGAGA 2201
QY 487 ValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPhe 506
DB 2202 GTCTTCAATGGTGGCCACATGTTCCATTTGACGTCCTCGTGAACCGCTTAAGTATGGTT 2261
QY 507 AsnArgThrIlePheGly 512
DB 2262 AACGAATGATCCACGGT 2279

```

Search completed: November 21, 2003, 20:21:28
 Job time : 133 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2003, 20:31:11 ; Search time 459 Seconds
(without alignments)
9809.723 Million cell updates/sec

Title: US-09-712-338-1
Perfect score: 1668
Sequence: 1 atcggtggctacgaattctt.....ccagtgtgtatggcatag 1668

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1662	99.6	1662	19	A. oryzae ATCC2038
2	894.6	53.6	1656	24	A. niger serine ca
3	313.6	18.8	3150	24	A. niger serine ca
4	164.8	9.9	1007	25	Aspergillus oryzae
5	134.4	8.1	1872	24	A. niger serine ca
6	118.6	7.1	1665	24	A. niger serine ca
7	116.8	7.0	1581	24	A. niger serine ca
8	116.8	7.0	2940	24	A. niger serine ca

9	114.6	6.9	3080	24	ABZ78226	A. niger serine ca
10	107.8	6.5	3221	24	ABZ78232	A. niger serine ca
11	103	6.2	678	21	AAFI2116	Aspergillus oryzae
12	87.8	5.3	636	21	AAFI3804	Aspergillus oryzae
13	86.4	5.2	2503	15	AAQ55347	Sequence of gene K
14	84	5.0	2002	17	AAQ55347	A. niger SFAG 2 ca
15	82.4	4.9	1611	24	ABZ78243	A. niger carboxype
16	78.2	4.7	2068	17	AAZ78283	A. niger carboxype
17	78.2	4.7	2660	24	ABZ78186	A. niger Bo-1 geno
18	75.6	4.5	626	21	AAFI2522	A. niger carboxype
19	68.4	4.1	1368	24	ABZ78267	Aspergillus oryzae
20	68.4	4.1	1653	24	ABZ78267	A. niger carboxype
21	68.4	4.1	2441	24	ABZ78210	C. albicans BAX-as
22	63.2	3.8	1446	24	ABZ78269	A. niger carboxype
23	63.2	3.8	2520	24	ABZ78212	A. niger carboxype
24	61.8	3.7	2027	24	ABZ78212	A. niger carboxype
25	58.6	3.5	1742	23	ABL07915	S. cerevisiae BAX-
26	58.6	3.5	3826	23	ABL07915	Drosophila melanog
27	58.6	3.5	4509	23	ABL08118	Drosophila melanog
28	57.4	3.4	475	25	ABZ54889	Aspergillus oryzae
29	56.4	3.4	1814	22	AAF68401	Human lung tumour
30	56.4	3.4	1814	24	ABK38312	cDNA encoding clon
31	56.4	3.4	1814	25	ACA10641	Human lung cancer-
32	56.4	3.4	1814	25	ABX99592	Lung cancer therap
33	56.4	3.4	1815	24	ABQ88162	Human osteoblast d
34	56.4	3.4	1815	24	ABN95798	Gene #2296 used to
35	54.8	3.3	2864	21	AAF18312	Lung cancer associ
36	49.2	2.9	1997	25	ACC46113	Human dithp enzyme
37	48.4	2.9	1839	24	ABZ78258	A. niger carboxype
38	48.4	2.9	2730	24	ABZ78301	A. niger carboxype
39	48	2.9	869	21	AAFI4932	Trichoderma reesei
40	47.6	2.9	390	13	AAQ21833	Randomising oligon
41	47.6	2.9	390	14	AAQ36859	PCR primer for 5'
42	47.6	2.9	390	22	AAF76910	Sequence containin
43	47.6	2.9	390	24	AAI72775	Oligo #7 for cloni
44	47.6	2.9	1428	22	AAE94477	Human hydrophobic
45	47.6	2.9	1551	19	AAV64076	Human serine carbo

ALIGNMENTS

RESULT 1

AAV28620
ID AAV28620 standard; DNA; 1662 BP.
XX
AC AAV28620;
XX
DT 27-AUG-1998 (first entry)
XX
DE A. oryzae ATCC20386 carboxypeptidase I DNA.
XX
KW Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous;
KW food industry; ss.
XX
OS Aspergillus oryzae.

XX
FH Key
FT CDS
FT Location/Qualifiers
FT 1..1662
FT /*tag= a
FT /product= carboxypeptidase I
FT /note= "partial coding sequence"

XX
PN W09814599-Al.
XX
PD 09-APR-1998.
XX
PF 03-OCT-1997; 97WO-US17977.
XX
PR 27-NOV-1996; 96US-0757534.
PR 04-OCT-1996; 96US-0726880.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.

ID ABZ78288 standard; cDNA; 1656 BP.
AC ABZ78288;
XX
DT 24-APR-2003 (first entry)
DE A. niger serine carboxypeptidase cDNA #3.
XX
KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW gene; ss.
XX
OS Aspergillus niger.
XX
FH Key Location/Qualifiers
FT CDS 1..1656
FT /*tag= a
FT /EC_number= "3.4.16.6"
XX
PN W0200268623-A2.
XX
PD 06-SEP-2002.
XX
XX 22-FEB-2002; 2002WC-EP01984.
XX
PR 23-FEB-2001; 2001EP-0200657.
PR 23-FEB-2001; 2001EP-0200658.
PR 23-FEB-2001; 2001EP-0200660.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.
PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-JUN-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.
PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 30-JUL-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.
PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000344.
PR 09-AUG-2001; 2001EP-0000357.
PR 16-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.
PR 22-OCT-2001; 2001EP-0000556.
PR 22-OCT-2001; 2001EP-0000557.
PR 22-OCT-2001; 2001EP-0000558.
PR 15-NOV-2001; 2001EP-0000464.
PR 21-DEC-2001; 2001EP-0000517.
XX
(STAM) DSM NV.
PA
XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI

PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albang R;
XX
DR WPI; 2002-723203/78.
DR P-PSDB; ABR38864.
XX
XX
PT Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications
XX
PS Claim 1; Page 252-253; 394pp; English.
XX
CC The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
CC A. niger proteases of the invention.
XX
SQ Sequence 1656 BP; 411 A; 426 C; 409 G; 410 T; 0 other;
Query Match 53.6%; Score 894.6; DB 24; Length 1656;
Best Local Similarity 71.3%; Pred. No. 1.6e-258;
Matches 1179; Conservative 0; Mismatches 474; Indels 0; Gaps 0;
QY 1 ATGCGTGGCTACGAATTTCTCTCAGTGTACCCCTGTTGTCAGCCAGTTGGGCCCTTCCA 60
Db |||||
QY 1 ATGCGTGGCTTCGGTTGGTCTTGTGGCCCTGGCTGACCTTAGTTGCTATGCC 60
Db |||||
QY 61 GGAGTACACCGCGCTCCGTGCGGTAGAGACAGCTACCAAGAACCCACCGGGTCAAG 120
Db |||||
QY 61 GAGATGAATGGTCACTACGATAGAGAGCGAGTTACCAAGACGCTCCACTGCGTCAA 120
Db |||||
QY 121 ACTCTTACACCGCAACCAATGTCACCATCCGGTACAGGAACCCCGGAGGGCGTC 180
Db |||||
QY 121 TCGATAAACCCCAACCAATGTCATGAGTATAGAGAACCAAGAACCCGAGGAATT 180
Db |||||
QY 181 TCGAGACTACCCCGGTGTCAAATCCTACTCTGGATATGTGCACACCTCTCCGAGTCC 240
Db |||||
QY 181 TGTGAGACAAACCTGGGGTCAAATCAATCACTCGGATATGTGATCTTTCGCCAGAGTCG 240
Db |||||
QY 241 CATACCTTTCTTGGTTCTTGAAGCCAGACATAACCCAGAACTGCACCTATCATTG 300
Db |||||
QY 241 CATACTTTCTTTGGTTTTCGAGTCAACCGCTGACCCGAAATGATCCAGTACTCTG 300
Db |||||
QY 301 TGGTTGAATGGGCGCTGGAAGCGATTTTGTATCGGTCTCTTGAAGAGTTGGCCCT 360
Db |||||
QY 301 TGGCTGAATGGGCGCTGGAAGCGATTTCTTGTATGGGCTTTTGAAGAGTTGGTCCG 360
Db |||||
QY 361 TGCCATGTCAATTCGACTTTTTCATGACTACATCAACCTCTCTGGAACAGGTCTCC 420
Db |||||
QY 361 TGTCAATCACACAGAGTACGAATCAATCAATCACTACTCTGGAACAGGTCTCC 420
Db |||||
QY 421 AATTACTATTCCTGTCCAGCCATTGGGAGTGGGCTTTTCATATAGTATACGGTTGAT 480
Db |||||
QY 421 AATCTTCTTTCTTGTCTCAGCCCTCGGTGGGGTCTCTTACAGTGAACAGGCC 480
Db |||||
QY 481 GGGTCCATTAACTGTAACTGGGGTGTGCGAAATTCGAGCTTTTCAGAGGAGTTGAGGC 540
Db |||||
QY 481 GGGTCTTGAATCCATTTTACTTGGAGCGGTGCGAAGACGGCTCTCTTGTGGAGTTT 540
Db |||||
QY 541 CGGTACCCAAACCATTTGATGCGCACTCTGATCGATATCTACCAATCTTGGCGAGGCCGT 600
Db |||||
QY 541 CGATACCCAGTTATTGATGCGCACTATCTATCGACACCGGATATCGTGCACGCGCAACC 600
Db |||||

PI Stiebler J, Albang R;
 XX WPI; 2002-723203/78.
 XX P-PSDB; ABR38864.
 XX
 PT Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications
 XX
 PS Claim 1; Page 187-189; 394pp; English.
 XX
 XX The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the
 CC A. niger proteases of the invention.
 XX
 SQ Sequence 3150 BP; 801 A; 778 C; 738 G; 833 T; 0 other;
 Query Match 18.8%; Score 313.6; DB 24; Length 3150;
 Best Local Similarity 72.5%; Pred. No. 2.3e-83;
 Matches 406; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
 QY 1091 ATGACCCGACTCGCCCAAGTTATTACAACAATTTCTGGCAAGGACTCTGTCAATGACG 1150
 Db 2194 AGGACCCGACCGCGCTCTACTTTGTGACTACCTCAAGAAGACTCAGTCATGGATG 2253
 QY 1151 CTATCGGCTCAACATCAACTACCCAGTCCAAATATGACGTTCTACTAGCTTTCCAGC 1210
 Db 2254 CTATCGGCTGGACATTAATCAACGAGTCCAGCGCGGAAGTATATTGATTCAGC 2313
 QY 1211 AAACGGCGACTTGTCTGGCCCACTTATCAAGACCTCGAGGAGATCCTTGTCTCC 1270
 Db 2314 AGACGGCGACTTGTGTGGCCGAATTTCAATGAGGACCTCGAAGAGATCCTCAACTCC 2373
 QY 1271 CCGTGGCTGTCTCCCTCACTATGCGGCGCGGATTCATCTGCACTGGTTCGGGGTC 1330
 Db 2374 CCGTACGGCTGTCTGATCTACGGCGATGCCGACTATATCTGTAATGTTGGCGGTC 2433
 QY 1331 AGCGGCTTCCCTCGTGCAGAACTACTCCAGCCGCCAGTTCCGAAAGCGAGGTACA 1390
 Db 2434 AGGCCATCTCACTCGCAGTTAACTACCCCATGTCAGTCACTGTTCCGTGAGCGGATCA 2493
 QY 1391 CCGCCCTGAAGTCAACGCGTTCGAGTATGGGAAACTCGGAGTATGTAATTTCTCT 1450
 Db 2494 CACCATGACAGTAGATGGGGTTCGATACGGTGAAGACTCGGAGTATGGCAACTTTTCT 2553
 QY 1451 TCACCTCGCTCTATGAGCGAGCGGCTGAAGTCCCATCTACAGCCCATCGCTCCCTGC 1510
 Db 2554 TCACCGGCTATATCAGGTGGCGCAGGAGTTCATATCACTATCAACGATCGCAGGTTGC 2613
 QY 1511 AATTGTTTAAACCGGACTATCTTTCGGTGGGATATCGCAGAGGCCCAAGAAGATCTGCG 1570
 Db 2614 AGCTGTTCAACGCTACTTTTATTTGGATGGATATTCAGCGGGTACAATCTCAGATTGGC 2673
 QY 1571 CCAGCTACAGAGGATGAACGGCTACAGTACGATACGATACAGTCTGCTCCGCTGC 1630
 Db 2674 CGGAATATAGCAACCAACGGGACATCGAGGCTACACACAGGAGTCTGTTGCGTCCACTGT 2733
 QY 1631 CTACGGCTACCAAGCATGTC 1650
 Db 2734 CCACGGCTCGAGTACCACC 2753

RESULT 4
 ABZ51997
 ID ABZ51997 standard; cDNA; 1007 BP.
 XX
 AC ABZ51997;
 XX
 DT 28-MAR-2003 (first entry)
 XX
 DE Aspergillus oryzae polynucleotide SEQ ID NO 1110.
 XX
 KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
 KW expressed sequence tag; gene; ss.
 XX
 OS Aspergillus oryzae.
 XX
 PN WO200279476-A1.
 XX
 PD 10-OCT-2002.
 XX
 PF 22-MAR-2002; 2002WO-1B00890.
 XX
 PR 30-MAR-2001; 2001JP-0098371.
 XX
 PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 PA (NARE-) NAT RES INST BREWING.
 PA (NORQ) NAT FOOD RES INST MIN AGRIC.
 XX
 PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
 PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
 XX
 WPI; 2003-046817/04.
 XX
 PT Detection of expression of specific Aspergillus genes for monitoring
 PT the fermentation and growth conditions of the fungus, using DNA probes
 PT
 XX
 PS Claim 1; SEQ ID NO 1110; 48pp + Sequence Listing; Japanese.
 XX
 CC The invention relates to a polynucleotide having any of 6006 specific
 CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
 CC specific culture conditions including one or more of eutrophic,
 CC oligotrophic, solid, early germination, alkaline, high temperature, low
 CC temperature or maltose culture or polynucleotides stringently hybridising
 CC to these sequences. The polynucleotides are useful for monitoring the
 CC progress of fermentation and the growth conditions of a fungus,
 CC especially of Aspergillus oryzae which is widely used in industrial
 CC fermentation. Also monitoring for fungal contamination.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1007 BP; 253 A; 270 C; 257 G; 224 T; 3 other;
 Query Match 9.9%; Score 164.8; DB 25; Length 1007;
 Best Local Similarity 52.4%; Pred. No. 8.1e-39;
 Matches 481; Conservative 3; Mismatches 390; Indels 44; Gaps 4;
 QY 684 GAGCTATGAGGCGCACTATGGTCTTCGCAATCTTCAATCATATTTTACGAGCAATGAGAG 743
 Db 42 GAGCTACGAGGACCGGTACGGCCCATCTTCTACTCTTCTTCAGGAGCAATGAGAA 101
 QY 744 AATTCGCAACGGTAGTGTTA-----ATGGTGTTCAGCTTAAATTCACACTCTCTGGGA 795
 Db 102 GATTGCCAACGCTCCATTGACATGACGATGCTACTATATTCACCTGACACCTCCGA 161
 QY 796 ATTATTAACGATCATCGACGAGCGGATCCAGGCCCTTACTACCTGAATTCGCTGTG 855
 Db 162 ATTATCAATGGCTGGCTCGATCTCCTGGTGGAGTCCCATCATATCCCCAGATAGCATAC 221
 QY 856 AACAAATACCTACGGTATCAAGGCTGTCAACGAGACCGCTCTACAACATCATGAGTTGCC 915
 Db 222 AACATACCTATGGCATTTGAGGCTATCAACAAACTGTATATGATATGATGCGGATGGAGCT 281

CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 CC A. niger proteases of the invention.
 XX
 XX
 SQ Sequence 1872 BP; 402 A; 434 C; 570 G; 466 T; 0 other;

Query Match 8.1%; Score 134.4; DB 24; Length 1872;
 Best Local Similarity 48.1%; Pred. No. 1.6e-29;
 Matches 644; Conservative 0; Mismatches 661; Indels 33; Gaps 8;

QY 240 CCATACCTCTCTCGTTCCTTTCGAGCCAGACAT-----AACCCAGAACTGCACCTAT 293
 DB 282 CAATACCTCTCTCGTTCCTTTCGAGCCAGACAT-----AACCCAGAACTGCACCTAT 341
 QY 294 CACATCTGTGTTGAATGTGGCCCTGGAAGCGATCTTTTGATCGCTCTCTTCGAGAGATT 353
 DB 342 CACCATCTGGATGAACCGCGGCGCGGGGATCTTCATGATTTGGGCTATTTCAGAGAA 401
 QY 354 GGGCCCTTGCATGTCAATTCGACTTTTGATGACTACATCAACCTCAC---TCGTGAA 410
 DB 402 CGGCGCACTGATCTGTAATACGGACTCGAATTCACGGCCCTAATCCCTGGTCTGTA 461
 QY 411 CGAGGCTCTCAATTTACTATTCCTGTCAGGCAATGGAGTCGGCTTTTCATATGTA 470
 DB 462 TGAGTAGCTGATATGTTGATATTTAGACAGCCGCTGACAGCGGATTTAGTTATGATGT 521
 QY 471 TAGCGTTGATGGGTCTCAATTAACCTGTAACTGGGGCTGTCGAAATTCGAGCTTTGAGG 530
 DB 522 GTTGAGGAATGGACGTTAGT-----TTGAATGAGACGTTTTTGGTGGGACGTTGCCG- 576
 QY 531 AGTTTACGGCCGTGACCAACCAATTCGATGCTCTGATCGATACCAATCTTCGCGC 590
 DB 577 -AGTCAGGATGTCATGGGACGGTGAATGGACGGTTAATGGGGAGGCGCTTTGGGT 635
 QY 591 AGAGGCGCTTGGAGATCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAG 650
 DB 636 TGGCTTGCAGTTTGGTGGGTGAATCTCTGAATATGTTCTTCTGTCACGCGAATGG 695
 QY 651 GGTGCTAGCTTAAGGACTTCACTCTATGGACGGAGAGCTATGGAGGCACTATGGTCTGC 710
 DB 696 TGGTGGTATGACAGGTTGATATATGGACGGAGTCATATGGGGAGCGGTATGGACCGGC 755
 QY 711 ATTCTTCAATCATTTTACAGACAGATGAGAGAAATGCAACCGTAGTGT---TAATGG 767
 DB 756 ATACACGGCGCTCTTTCAGGAGATGAATGAGAGGATGAGATGGGGAGTAGACACCGG 815
 QY 768 TGTTCAGCTTAATTTCAACTCTCTGGGAATTTATACGGCATCATCGAGCGCATCCA 827
 DB 816 GAAGAAGATCCATTTGGATACGCTGGGCAATTAATGAGTGGGTGTGATCTCGTGCA 875
 QY 828 GGCCTCTTACTACCTGAAATTCGCTGTGAACATACCTACGCTATCAAGCTGTCAACGA 887
 DB 876 GTTCCTCTGTTCCCTGACAGCGGTATTAACATATGATGAGATCGAGGGAATCAATCG 935
 QY 888 GACCGTCTAACAATACATGAAGTTTCCCAACCAATGCTGCAAGATGCTGATTTGAT 947
 DB 936 CACGCTCTACGACCGGCTATGATAGTTGGAGACGCTTGGCGGTGAGGGATATGAT 995
 QY 948 TTCCACCTTCAACACAGACAAACCGCACCGATAGCTGACTACGCTCTGCGCGAAGC 1007
 DB 996 CATCGAGTCTCGGATGCTTGGGAGCTCGG---AGATCCCTCATCATCTGCGAGAGCG 1052
 QY 1008 CACCAACATGTGAGGACAAATTTGAGGGGCACTACTACGCTTTGCTGTTGCTGTGT 1067
 DB 1053 GTCGGACTACTGTTTCGGGAGATCAGAGCTGTATACGAATACCTCCGCGGAGGATA 1112
 QY 1069 GTATGATTTCCGATTCATATGATGACCGGACTCGCCCAAGTTATTAACAATAATTTCT 1127
 DB 1113 CTACGACATAGGCAATTTTCAACCGGATGACGCTCTCTGTCCTTACTTCTGTCGGGTTCT 1172
 QY 1128 GCAAGAGGACTCTGTCATGAGGCTATCGGCGTCAACATCAACTACCCAGTCCAATAA 1187

DB 1173 GAATCGCCATGGGTGCAAAAGGCACTTGGGGTCCCGTGAACTATACCATGTCTGTCAGA 1232
 QY 1188 TGAGCTCTACTACGCTTTTCCAGCAAAACAGGCGACTTTGCTGTGGCCCACTTTCATCGAAGA 1247
 DB 1233 GGCAGTGGGAAACAGTTTCGCTCGACGGGCGAATATCCGCGAAATGATCCCGCGGAAT 1292
 QY 1248 CCTCGAGGAGA-----TCCTTGTCTCTCCCGTGGGTGCTCCCTCACTATATGGCGA 1298
 DB 1293 GATCGGGGATATTGATATCTTGTGACTCCGCTGTCAAGGTGGCTATGTTATATGGGA 1352
 QY 1299 CGCCGATTTACATCTGCAACTGTTTCGGGTCAGGCGGTTTCCCTCGCTCGCTGCAACTACTC 1358
 DB 1353 CCGGACTATGCTTGTTCGGTGGCGCGGGAAGATGTCAAGCTGTGTTGGTGGAGTACGA 1412
 QY 1359 CCAAGCCGCCAGTTCCGAAAGCGAGGATACACGCCCTGAAAGTCAACGCGCTCGAGTA 1418
 DB 1413 GGATCGGAGAGTTCCGTTGCTGCTGGTATGCCGAGTGCAGAGAGTCACTCTAGT 1472
 QY 1419 TGGGG---AACTCGGAGTATGTAATTTCTCTTCTCCTGCTGCTGATGAGGCGGCCA 1475
 DB 1473 TGGGGGCTTAGTAAGGAGTATGGAACCTTCTGTTTCCGCTGTCTTTCAGGCGGCCA 1532
 QY 1476 TGAAGTCCCATCTACTACAGCCCATCGCTCCCTGCATTTGTTAACCGGACTATCTTCGG 1535
 DB 1533 TGAGTCCCATTTTATCAGCCCGAAACGCGGTATGAGATTTTAAATCGCGCTCAGTTTAA 1592
 QY 1536 TTGGGATATCCGAGAGGG 1553
 DB 1593 TTGGGATATTGCGACGGG 1610

RESULT 6
 ABZ78283
 ID ABZ78283 standard; cDNA; 1665 BP.
 XX
 AC ABZ78283;
 XX
 DT 24-APR-2003 (first entry)
 XX
 DE A. niger serine carboxypeptidase cDNA #2.
 XX
 KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ss.
 XX
 OS Aspergillus niger.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1665
 FT /*tag= a
 FT /EC_number= "3.4.16.6"
 XX
 PN WO200268623-A2.
 XX
 PD 06-SEP-2002.
 XX
 PF 22-FEB-2002; 2002WO-EP01984.
 XX
 PR 23-FEB-2001; 2001EP-0200657.
 PR 23-FEB-2001; 2001EP-0200658.
 PR 23-FEB-2001; 2001EP-0200660.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 26-FEB-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0000075.
 PR 28-MAR-2001; 2001EP-0000078.
 PR 28-MAR-2001; 2001EP-0000080.
 PR 28-MAR-2001; 2001EP-0000087.
 PR 28-MAR-2001; 2001EP-0000088.
 PR 21-MAY-2001; 2001EP-0000156.
 PR 21-MAY-2001; 2001EP-0000159.

PR	21-MAY-2001; 2001EP-0000160.	Db	190	GCCTCAACAGCTCTCTTTTGGTTTTCGAAGCCCGCAAGATCCAGCAATGGCGCTCTG	249
PR	21-MAY-2001; 2001EP-0000162.	QY	295	ACATTGGTGAATGGTGGCCCTGGGAAGCAATCTTTGATCGGTCTCTTTGAAGAGTTG	354
PR	21-MAY-2001; 2001EP-0000165.	Db	250	GCCATCTGGCTCAATGGGGTCCGGGTGGCTCGCTCATGGGCTCTCTTTGAAGAAATTA	309
PR	21-MAY-2001; 2001EP-0000168.	QY	355	GGCCCTTGCATGTCATTTGACATTTTGTATGAC---TACATCAACCCCTCACTCTGGGAAC	411
PR	21-JUN-2001; 2001EP-0000240.	Db	310	GGTCCCTTGTTCATTCGATCAGACTCCAGAGCAAGTCCCTCAATCTTGGATTTGGAAC	369
PR	21-JUN-2001; 2001EP-0000242.	QY	412	GAGGTCTCCAAATTTACTATTCTGTCCAGCCATTGGGAGTCGGCTTTTCAATATAGTAT	471
PR	21-JUN-2001; 2001EP-0000244.	Db	370	AATGAAGTCAATCTCTATTCTTTCACCAAGCACTCAAGTCGGCTTCTCATACGATGTC	429
PR	21-JUN-2001; 2001EP-0000246.	QY	472	ACGTTGATGGTCCATTAACCCCTGTAACTGGGGTCTCGAATAATTCGAGCTTTCAGGA	531
PR	21-JUN-2001; 2001EP-0000248.	Db	430	CCACAAATGGCACTTTGACAGCTA-----	454
PR	21-JUN-2001; 2001EP-0000250.	QY	532	GTTCAGGGCCGTATCCCAACCAATTCATGCCACTCTGTATGATCACTACCAATCTTGC	591
PR	21-JUN-2001; 2001EP-0000252.	Db	455	-----ATGGGACTGCATTCGGGCT	474
PR	21-JUN-2001; 2001EP-0000254.	QY	592	GAGGCGCTTGGGAGATCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTTAGG	651
PR	21-JUN-2001; 2001EP-0000256.	Db	475	CRGCTCTATGCAATTCGCGCAACCTGGTTTTTCGAGTTCCTCACTACAAGCCAAAC	534
PR	21-JUN-2001; 2001EP-0000258.	QY	652	GTGAGTCTAAGGACTTCACTCTATGACGGAGAGCTATGAGGGGCACTATGGTCTGCA	711
PR	21-JUN-2001; 2001EP-0000260.	Db	535	GATGATCGTG-----TCAGTCTCTGGCTGAAAGTTACGGAGGCAATTTATGGTCCAGG	588
PR	21-JUN-2001; 2001EP-0000262.	QY	712	TTCTTCAATCAATTTTACGAGCAGATGAGAAATTTGCCAAGCTAGTGTATATGGTGT	771
PR	21-JUN-2001; 2001EP-0000264.	Db	589	ATCTTTCGGTCTTCCACAGCAGAAATGACAAATTCGAGAGGGGACTGCAGAGACGGT	648
PR	21-JUN-2001; 2001EP-0000266.	QY	772	-----CAGCTTAATTTCAACTCTCTGGGAAATTTATTACGGCATCATCGAGAGGCA	825
PR	21-JUN-2001; 2001EP-0000268.	Db	649	GCACAGTATTTGCATCTCGACAGCTTGGCAATTTGTGAACGGCTTGATGATATGGTATC	708
PR	21-JUN-2001; 2001EP-0000270.	QY	826	CAGGCGCTTACTACCTGAAATTCGCTGTGAACAATACCTACGATCAAGGCTGTCAC	885
PR	21-JUN-2001; 2001EP-0000272.	Db	709	CAAGAAGAGGCTTTACATTTACTTGGCCATACAATA--ACGTAAAGGCTCGCCCTCTCAT	766
PR	21-JUN-2001; 2001EP-0000274.	QY	886	GAGACGCTCTACAACCTACATGAAGTTTGCACCAAAATGCAATGGTTCCAGGATTTG	945
PR	21-JUN-2001; 2001EP-0000276.	Db	767	TCAACTCGCGAGGCTTTCGGATCAGGCGCTCGCTCGGAGCGGCTTTGAAGAGCGG	826
PR	21-JUN-2001; 2001EP-0000278.	QY	946	ATTTCCACCTGCAACAGACAAACCCGCAATTAGCTGACTACGCTCTCGCGCGAA	1005
PR	21-JUN-2001; 2001EP-0000280.	Db	827	ATTTCCGGCTTGC-----TCACTCAGGGAAGAAATATCTCTGAAATTTGCGGA	873
PR	21-JUN-2001; 2001EP-0000282.	QY	1006	GCACCAACATGTGCGGAGCAATGTGAGGGCCATCTACGCCCTTGTGCTGCTGCT	1065
PR	21-JUN-2001; 2001EP-0000284.	Db	874	GGCTTGCACATAGATGGGAGATGGCCCATCACCTACTACCACTTCAATTCGCGGG	933
PR	21-JUN-2001; 2001EP-0000286.	QY	1066	GTGTATCATATTCGGATCCATATGATGACCCGCTCCGCAAGTTATTATCAACAATTT	1125
PR	21-JUN-2001; 2001EP-0000288.	Db	934	TGTATCAGCATTCGCCCATCTTAAGACAGCCATTCCTCCCAAGCAGATGCTCGATAT	993
PR	21-JUN-2001; 2001EP-0000290.	QY	1126	CTGGCAAGAGCACTCTCTATGAGCGCTATCGGCGTCAACATCACTACACCCAGTCCAAT	1185
PR	21-JUN-2001; 2001EP-0000292.	Db	994	TTGAGCGAGGAGTCCGTCCTTGGCGCTCTTGGGCTACAGTCAATTTTCAATGCTCTCG	1053
PR	21-JUN-2001; 2001EP-0000294.	QY	1186	AATGAGCTCTACTACGCTTTCAGCAAGAGGCGACTTCTCTGGCCCACTTCAATCGAA	1245
PR	21-JUN-2001; 2001EP-0000296.	Db	1054	AGTGGCGGTGCTACAGATTTCAATAAAACCTTTGATATCGTCCAGCGGCTTCTCGAT	1113
PR	21-JUN-2001; 2001EP-0000298.	QY	1246	GACCTCGAGGAGATCTCTGCTCTCCCGTGGTGTCTTCCCTCATCTATGCGAGCGCGAT	1305
PR	21-JUN-2001; 2001EP-0000300.	Db	1114	GCAATTTGGCTACCTCTCGACAGTGGTGTAAAGTACACATGATGTACGGAGATCGTAT	1173
PR	21-JUN-2001; 2001EP-0000302.	QY	1306	TACATCTGCAACTGTTTCGGCGGTTCAGCGGCTTTCCTCTCGTGGAACTACTCCCAAGCC	1365

(STAM) DSM NV.

Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
Stiebler J, Albang R;
WPI; 2002-723203/78.
P-PSDB; ABR38859.

Novel isolated protease polypeptide useful in laboratory, clinical, pharmaceutical, chemical, diagnostic, personal care and industrial applications

Claim 1; Page 246-247; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A polypeptide or polynucleotide of the invention is useful for diagnosing a fungal infection such as aspergillosis, or as a query sequence to perform a search against public databases. A polypeptide of the invention is useful in a selected number of industrial or pharmaceutical processes, in laboratory or clinical processes, in food industry (baking, brewing, cheese manufacturing, meat tenderising), in tanning industry and in the manufacture of biological detergents. A polypeptide may also be useful for improving protein solubility, extraction yields, viscosity or taste, texture, nutritional value, minimising of antigenicity or anti-nutritional factors, colour or functionality as well as processing aspects like filterability of the proteinaceous raw material. The sequences shown in AB278237-AB278293 represent cDNA encoding the A. niger proteases of the invention.

Sequence 1665 BP; 374 A; 455 C; 424 G; 412 T; 0 other;

Query Match 7.1%; Score 118.6; DB 24; Length 1665;
Best Local Similarity 47.2%; Pred. No. 8.8e-25;
Matches 673; Conservative 0; Mismatches 649; Indels 105; Gaps 6;

QY 175 GGCTCTGCGAGATACCCGGGTGTCAAATCTCTCTGATATGTCGACACTCTCC 234
|||
Db 130 GGAATTTGCGAACTACCGGGTGTCCGATCTTATTGGGCTATGACACTTCC 189
|||

QY 235 GAGTCCCATACCTCTCTGCTTCTTCGAAGCCAGACATACCCAGAACTGCACCTATC 294

•

PR 12-

QY 1208 AGCAACAGGCGACTTTCTGTGGCC---AACTTTCATGAAGACCTCGAGAGATCCTTG 1264
 Db 1732 CGTCCACCGGTGACGACGGCGACTCATGAACACCATCCAGATCTGGGAGCTGCTCA 1791
 QY 1265 CTCTCCCGTGGTCTCTCCCTCATCTATGCGAGCGCGATTACATCTGCAACTGGTTCG 1324
 Db 1792 AACAGGTGTACGGTGTGTATGACCGCGGGATGCCGACTATAAATGCAACTGGCTGG 1851
 QY 1325 GCGTCAAGCGGCTTTCCCTCGCTCGAACTACTCCCAAGCGCCAGTTCGGAAGCGCAG 1384
 Db 1852 GTGGGGAAGCGTGT-----CGTTGAGGTCAAGCGCGCCCAACTTCAGTAGTCGG 1902
 QY 1385 GTACACGCCCC---CTGAAGTCAACGCGTCTGAGTATGGGAAACTCGCGAGTATGTA 1441
 Db 1903 GTTACACCAACATTTGTACCTCGGATGGAGTGACACAGCGCCAGTGGCGCGCGGGC 1962
 QY 1442 ATTTCTCTTCACTCGGCTCATGAGGAGCGCCATGAATGCCATATACAGGCCATTCG 1501
 Db 1963 AATTGCTTTGTGCGAGTGTATGAGAGTGACATGAGTTCCCTTCTATCAACCTTGC 2022
 QY 1502 CTCCTCGCAATTTTAAACGGAATCTCTCGGTTGGGATATCGCAGAGGGCCAGAGA 1561
 Db 2023 TTGGCTGGAGATTTTGAAGCGCTCATTTGGCGGCAAGGATGTGGCAGGGAAGATTC 2082
 QY 1562 AGATCTGCGCCAGCTCAAGACGAATCGAAGCGGCTA 1597
 Db 2083 CCATCTCGTGGATTACAGACGGTGGGCGACGCCCA 2118

RESULT 9

ABZ78226

ID ABZ78226 standard; DNA; 3080 BP.

XX XX

AC ABZ78226;

DT 24-APR-2003 (first entry)

XX XX

DE A. niger serine carboxypeptidase gene #2.

XX XX

KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ds.

XX XX

OS *Aspergillus niger*.

XX XX

PN WO200268623-A2.

XX XX

PD 06-SEP-2002.

XX XX

PF 22-FEB-2002; 2002WO-EP01984.

XX XX

PR 23-FEB-2001; 2001EP-0200657.

XX XX

PR 23-FEB-2001; 2001EP-0200658.

XX XX

PR 23-FEB-2001; 2001EP-0200660.

XX XX

PR 26-FEB-2001; 2001EP-0200706.

XX XX

PR 26-FEB-2001; 2001EP-0200707.

XX XX

PR 26-FEB-2001; 2001EP-0200719.

XX XX

PR 26-MAR-2001; 2001EP-0000075.

XX XX

PR 28-MAR-2001; 2001EP-0000078.

XX XX

PR 28-MAR-2001; 2001EP-0000080.

XX XX

PR 28-MAR-2001; 2001EP-0000087.

XX XX

PR 21-MAY-2001; 2001EP-0000088.

XX XX

PR 21-MAY-2001; 2001EP-0000156.

XX XX

PR 21-MAY-2001; 2001EP-0000159.

XX XX

PR 21-MAY-2001; 2001EP-0000160.

XX XX

PR 21-MAY-2001; 2001EP-0000162.

XX XX

PR 21-MAY-2001; 2001EP-0000165.

XX XX

PR 21-MAY-2001; 2001EP-0000166.

XX XX

PR 21-MAY-2001; 2001EP-0000168.

XX XX

PR 21-JUN-2001; 2001EP-0000240.

XX XX

PR 21-JUN-2001; 2001EP-0000242.

XX XX

PR 21-JUN-2001; 2001EP-0000244.
 PR 21-JUN-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 12-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000323.
 PR 30-JUL-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 22-OCT-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0004464.
 PR 21-DEC-2001; 2001EP-0005117.
 XX XX

(STAM) DSM NV.

Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 Stiebler J, Albarg R;

MPI; 2002-723203/78.

P-PSDB; ABR38859.

Novel isolated protease polypeptide useful in laboratory, clinical,
 pharmaceutical, chemical, diagnostic, personal care and industrial
 applications

Claim 1; Page 178-180; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A
 polypeptide or polynucleotide of the invention is useful for diagnosing a
 fungal infection such as aspergillosis, or as a query sequence to perform
 a search against public databases. A polypeptide of the invention is
 useful in a selected number of industrial or pharmaceutical processes, in
 laboratory or clinical processes, in food industry (baking, brewing,
 cheese manufacturing, meat tenderising), in tanning industry and in the
 manufacture of biological detergents. A polypeptide may also be useful
 for improving protein solubility, extraction yields, viscosity or taste,
 texture, nutritional value, minimising of antigenicity or
 anti-nutritional factors, colour or functionality as well as processing
 aspects like filterability of the proteinaceous raw material. The
 sequences shown in ABZ78180-ABZ78236 represent genes encoding the
 A. niger proteases of the invention.

Sequence 3080 BP; 755 A; 803 C; 735 G; 787 T; 0 other;

Query Match

Best Local Similarity 6.9%; Score 114.6; DB 24; Length 3080;

Matches 279; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

QY 1040 CATACTACGCTTGTCTGTGTGTGTATGATTTGGCATCATATGACCCGA 1099
 Db 1771 CCTACTACCACTTCAATCGCGGTGTGTACGATCGCCATCTTAAGACGCCAT 1830
 QY 1100 CTCGCCCAAGTTATTACAAATTTCTGGCAAGACTCTGTCTGACGCTATCGCG 1159
 Db 1831 TCCCTGCCAGCACATGCTCGGATATTTGACCCAGAGTCCGTCTTGGCTCTTGGGG 1890
 QY 1160 TCAACATCAACTACACCCAGTCCAAATAGCTCTACTACGCTTTCCAGCAACAGGG 1219
 Db 1891 TACCAGTCAATTCACATCGTCTTCGAGTGGCGGTGCTACACAGTTCATAAACCCTTTG 1950

QY 1220 ACTTGTCTGCCCCAACTTCAATCGAAGACCTCGAGGAGATCCTTCTCTCCCGGTGCGT 1279
Db 1951 ATATCGTCCACGGCGCTTCTGGATGCAATTTGGCTACTCTCGACAGTGGTGTAAG 2010
QY 1280 TCTCCCTCATCTATGGACGCGGATTAATCATCTGCACTGTTTGGCGGTGAGGCGGTTT 1339
Db 2011 TACACATGATGTACGAGATCGTGATTAATGCTGCAATTTGGGTGCGGGGCGAAGGCA 2070
QY 1340 CCTCGCTGCGAATCTCCCAAGCGCGCCAGTTCGGAAGCGCAGGGTACACGCCCTTGA 1399
Db 2071 GCCTTGAGTTCGATTTCCCGTATACCGAATTTGCCGACACGGGATCTCCCACTCC 2130
QY 1400 AAGTCAACGGGTGAGTATGGGAACTCGGAGTATGTTAAATTTCTCTTCACTCGG 1459
Db 2131 TTACGCCCGACGGGATCAGCGCATGACCGCCAGCTGGGCAACTACAGCTTCACTCGG 2190
QY 1460 TCTATGAGCGGCGCATGAGTCCCATCTACTACGCGCCATCGCTCCCTGCAATTTGTTA 1519
Db 2191 TCTTCAAGCGGGGATGAGGTCCCTCTACAGCTGTCGGCGGTATGAGATCTTCA 2250
QY 1520 ACCGACTATCTTCGTTGGGATATCGCAGAGGGGCGAGAAGAGATCTGGCCAGCTACA 1579
Db 2251 TCGGGCGCATCTCAACAGATATCCCTACTGGCTCTTGGCTGTTGATGACGAATTC 2310
QY 1580 AGACGAATGGAAC 1592
Db 2311 AGTCGTTGGACC 2323
RESULT 10
ID ABZ78232 standard; DNA; 3221 BP.
XX AC ABZ78232;
XX DT 24-APR-2003 (first entry)
DE A. niger serine carboxypeptidase gene #4.
XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
XX Gene; ds.
XX OS Aspergillus niger.
XX PN WO200268623-A2.
XX FD 06-SEP-2002.
XX PF 22-FEB-2002; 2002WO-EP01984.
XX PR 23-FEB-2001; 2001EP-0200657.
PR 23-FEB-2001; 2001EP-0200658.
PR 23-FEB-2001; 2001EP-0200660.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.
PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-JUN-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.

PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 30-JUL-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.
PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000357.
PR 09-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.
PR 22-OCT-2001; 2001EP-0000556.
PR 22-OCT-2001; 2001EP-0000557.
PR 22-OCT-2001; 2001EP-0000558.
PR 15-NOV-2001; 2001EP-0004464.
PR 21-DEC-2001; 2001EP-0005117.
XX (STAM) DSM NV.
XX PA Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI Klugbauer S, Wagner C, Fritz A, von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albang R;
XX WPI: 2002-723203/78.
DR P-PSDB; ABR38865.
XX PT Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications
XX PS Claim 1; Page 189-191; 394pp; English.
XX CC The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the
CC A. niger proteases of the invention.
XX SQ Sequence 3221 BP; 739 A; 773 C; 869 G; 840 T; 0 other;
Query Match 6.5%; Score 107.8; DB 24; Length 3221;
Best Local Similarity 48.5%; Pred. No. 2.2e-21;
Matches 449; Conservative 0; Mismatches 437; Indels 39; Gaps 4;
QY 668 TCAGTCTATGACGGAGAGCTATGAGGAGGACATATGCTCTGCTTCAATCAATTTT 727
Db 1522 TGAGTATGACGGAGTATATGAGGAGGAGTATGACGGGATACACGGGCTCTTTC 1581
QY 728 ACCGAGCAATGAGAAATTCACCGGATGCTTAAAT---GGTGTTCAGCTTAATTC 784
Db 1582 AGGAGATGAATGAGAGGATGAGTGGGAGGTAAGACCGGGAAGAGATCCATTTGG 1641
QY 785 ACTCTCTGGGAATTAATTAACGGCATCATCGACGAGGATCCAGGCCCTTACTACCTG 844
Db 1642 ATACGCTGGGCAATTAATCAATGGGTGTGGATTACTCGTGAGGTCCCTTCCTG 1701
QY 845 AATTGCTGTGAACAATACCTACCGTATCAAGGCTGTCAACGAGACCGTCTACAACTACA 904

Db 1702 AGCAGCGGTATACCAATAGTATGGATCGAGGGAATCAATCGCAGCTCTACGACGGG 1761
 QY 905 TGAAGTTTGCACCAATGCTTCCAGGATTTGATTTCACTG----- 956
 Db 1762 CTATGGATGTTGAGCAGGCTCGGGTGCAGGATATGATCATGAGTGCAGGATG 1821
 QY 957 -----CAACAGACAAACCGCATTTAGCTGACTACGCCCTCTGCG 1000
 Db 1822 CTGCGAGCTCGAGATCCCTCTATGATGCGACATGAGACGTAATAGCATCTGCG 1881
 QY 1001 CCGAAGCCACCAATGTCAGGACAAATGTTAGGGGCCATATAGCCCTTTCTGCTC 1060
 Db 1882 AGGAGGCTCGGATCTACTGTTCCGCGGAGATCAAGAGCTGTTATACGAATACCTCCGGC 1941
 QY 1061 GTGTGTGTATGATATTCGGCATCCATATGATGACCGCATCCCGCAAGTTATTACACA 1120
 Db 1942 GAGGATACGACATAGCGCATTTTCAGCGCGATGAGCTCTCGTCTTACTTCTGCG 2001
 QY 1121 AATTTCTGGCAAGGACTGTGTATGACGCTATFCGGGTCAACATCAACTACACCCAGT 1180
 Db 2002 GGTCTTGAATCGCCATGGTGCAGAGGCACTTGGGTCCCGGTGCACTATACCATGT 2061
 QY 1181 CCAATATGAGCTTACTAGCTTTCCAGCAACAGCGGACT-----TTGTCTGCG 1231
 Db 2062 CGTCAGAGCGAGTGGGAAAGTTCGCTCGACGGCGGATTTCCGCGAAATGATCCOC 2121
 QY 1232 CCAACTTCATCGAAGACTCGAGGAGATCTTGTCTCTCCCGTGGGTCTCCCTCATCT 1291
 Db 2122 GCGGAATGATCGGGATTTGGATCTTGTGACTCCGGTGTCAAGGTGGCTATGAT 2181
 QY 1292 ATGGGACCGGACTATGCTTGTCCGTTGGCGGTGAGCGGCGGAGATGTCAGCTCTGCTG 1351
 Db 2182 ATGGGACCGGACTATGCTTGTCCGTTGGCGGTGAGCGGCGGAGATGTCAGCTCTGCTG 2241
 QY 1352 ACTACTCCCAAGCGCCATTCGAGCCAGGATACAGCCCTTGAAGTCAACGGCG 1411
 Db 2242 AGTACGAGATGCGGAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2301
 QY 1412 TCAGTATGCGG---AACTCGCGAGTATGATGATTTCTCTTCACTCGGCTATGAGG 1468
 Db 2302 CCTACGTTGGGTCTAGTAAAGCATGATGGAACTTCTCTGTTCAAGCGGTCTTTT 2361
 QY 1469 CAGGCATGAGTCCATACCTACAGCCATCGCTCCCTGCAATGTTTAAACGGGACTA 1528
 Db 2362 CCGGCCATGAGTGCCATTTATCAGCCGGAACGGCGATGAGATTTTAACTCGGCTC 2421
 QY 1529 TCTTCGTTGGGATTCGACAGGG 1553
 Db 2422 AGTTTAATGGGATATTCGACGGG 2446

RESULT 11
 AAF12116

ID AAF12116 standard; cDNA; 678 BP.

XX AAF12116;
 AC AAF12116;

XX 13-MAR-2001 (first entry)
 XX Aspergillus oryzae EST SEQ ID NO:4639.

XX Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX Aspergillus oryzae.

OS WO200056762-A2.

XX 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.
 XX 22-MAR-1999; 99US-0273623.
 PR (NOVO) NOVO NORDISK BIOTECH INC.
 XX (NOVO) NOVO NORDISK AS.
 PA Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI; 2000-594572/56.
 DR Monitoring differential expression of genes in filamentous fungal cells
 XX uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 PT Claim 88; Page 1971; 316pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

SQ Sequence 678 BP; 153 A; 188 C; 175 G; 161 T; 1 other;

Query Match 6.2%; Score 103; DB 21; Length 678;
 Best Local Similarity 53.0%; Pred. No. 2.7e-20;

Matches 273; Conservative 0; Mismatches 230; Indels 12; Gaps 2;

QY 1055 CTGGTCTGGTGTATGATATTCGGCATCCATATGATACCGGACTCCGCCAAGTTATT 1114
 Db 3 CTGGACGAGGTATTACTACATCTCGCATTTTGACCCCATCTTTCCCCCTCCGACT 62

QY 1115 ACAACAAATTTCTGGCAAGGACTCTGTCATGACGCTATCGGCTCAACATCACTACA 1174
 Db 63 ACTTTGGTTCTTAACACGACCTGGTCCAAAGGAGCCCTCGGGTTCGGTTAAATTCA 122

QY 1175 CCAGTCCCAATATGACGCTTACTACTGCTTCCAGCAACAGCGGACTTTGCTGCCCCA 1234
 Db 123 CGGAATCTGTTGATAGCGTCTACATGCTTTTCGCAACAGTGACTACCCACGCTCG 182

QY 1235 ACTTCATCGAGACTCTCGAGGATC-----CTTGCTCTCCCGTGGTGTCTCC 1285
 Db 183 ATGTACGGGGTACCTCGGAGATATCGCATACGCTTGACTCTGGCATCAAGTTGCC 242

QY 1286 TCATCTATGGGACCCCGATTCATCTGCACTGTTGGCGGTGAGGCCGTTCCCTCG 1345
 Db 243 TAGTGTATGGCGATCGGATTTACGATCCCTCGAAGGAGAGAGTGAAGTGA 302

QY 1346 CTGCGAATCTATCCCAAGCCGCCAGTTCCGAAGCGAGGTACACGCCCTCGAAGTCA 1405
 Db 303 AGGTGAGTACTCGATGCGCCCAAGTTCCGCTCTGCGGTTACGCCCTCTGAAGACCA 362

CC niger; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 636 BP; 148 A; 177 C; 155 G; 156 T; 0 other;

Query Match 5.3%; Score 87.8; DB 21; Length 636;
Best Local Similarity 54.9%; Pred. No. 9.9e-16;
Matches 195; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

QY 174 GGGGCTCTGCGAGACTACCCCGGTGTCAAAATCCTACTCTCGATATGTCGAC---ACCTC 230
Db 79 GAGGCTCTCGGTATCGACCCGAAATGTGAAGCAATACACTGTTATCTCGACGATAACGG 138
QY 231 TCCGAGTCCCATACCTTCTTCTGTTCTTTCGAGCCAGACATACCCAGAAACTGCACC 290
Db 139 TAATGACAAAGCATCTGTTCTACTGTTCTTCTGAAATCTCGCAATGATCCTTAAGAACGACCC 198
QY 291 TATCACATTGTGTTGAATGTGGCCCTGGAAGGATTCTTTGATCGGTCTCTTCGAAGA 350
Db 199 CGTCGTCTCTGGCTGAATGGGCGCGGTGCTCTTCCCTCACTGGTCTATTATGA 258
QY 351 GTTGGGCGCTTGCATGTCAATTCGACTTTTGTATGATACATCAACCCCTCACTCGTGGAA 410
Db 259 GCTGGACCTAGCAGCATGAGACATTAAGCCCGTCTACAATGACTTCTCTGTTGAA 318
QY 411 CGAGGTCTCCAAATTTACTTCTCTGTCAGCCATGGAGTCGGCTTTTCATATAGTA 470
Db 319 CTCTAACGCTCCGTCATTTCTCTTGTATGACGCCCGTGAACGTCGGCTATTCCTACAGTGG 378
QY 471 TACGGTTGATGGTCCATTAAACCTGTAACCTGGGTCTGCGAAGAAATTCGAGCTTT 525
Db 379 CTCGCTGTATAGCAGACTGTTGCTGCTGCGAAGGATGCTAGGCTCTCTCTCT 433

RESULT 13
AAQ55347
ID AAQ55347 standard; cDNA; 2503 BP.
XX
AC AAQ55347;
AC AC
XX 25-MAR-2003 (updated)
DT 20-JUL-1994 (first entry)
XX
DE Sequence of gene Kl.PRC1 encoding protease C.
XX
KW Protease; yeast; proteolysis; ss.
XX Kluyveromyces lactis.
OS
XX
FH Key Location/Qualifiers
FT CDS 387..1862
FT /*tag= a
XX
XX WO9400579-A1.
XX
XX 06-JAN-1994.
XX
XX 23-JUN-1993; 93WO-FR00623.
XX
XX 25-JUN-1992; 92FR-0007785.
XX
XX (RHON) RHONE POULENC RORER SA.
XX
XX Fleer R, Fournier A, Yeh P;
XX WPI; 1994-026215/03.
XX P-PSDB; AAR48059.
XX
XX New Kluyveromyces yeast with modified protease gene - esp. used
XX for high yield prodn. of recombinant protein, also DNA encoding
XX yeast protease and derived peptide(s)
XX

QY 1406 ACGCGTCGATATGGGAAACT---CGGAGTAGTGGTAATTTCTCTTCACTCGCGTCT 1462
Db 363 ATGCTCATATGTTAGTGGCTTGTGTCGACAGTACGGAAACTTCTGTTCACTCGCGTCT 422
QY 1463 ATGAGGAGCGCATGAAGTCCATCTACAGCCCATCGCTCTCCCTGCAATGTTTAAAC 1522
Db 423 TCGAAGCGGTCATGAGGTGCGGCATATCATGACCTGAAACGCGGTATGAGATCTTCCACC 482
QY 1523 GGACTATCTCGGTTGGGATATCGCAGAGGGCCAG 1557
Db 483 GAGCATTTGTTCAACAGACATTTGCGACGGGCAAG 517

RESULT 12
AAFL13804
ID AAFL13804 standard; cDNA; 636 BP.
XX
AC AAFL13804;
AC AC
XX 13-MAR-2001 (first entry)
DT
DE
DE Aspergillus oryzae EST SEQ ID NO:6327.
XX
XX Multiple gene expression; filamentous fungal cell; EST;
XX expressed sequence tag; *Fusarium venenatum*; *Aspergillus niger*;
KW *Aspergillus oryzae*; *Trichoderma reesei*; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS
XX
XX *Aspergillus oryzae*.
XX
XX WO2000056762-A2.
XX
XX 28-SEP-2000.
XX
XX 22-MAR-2000; 2000WO-US07781.
XX
XX 22-MAR-1999; 99US-0273623.
XX
XX (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags -
XX
XX Claim 88; Page 2600; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring
XX the global expression of genes from FF cells allows the production
XX potential of the microorganisms to be improved. New genes may be
XX discovered, possible functions of unknown open reading frames can be
XX identified and gene copy number variation and stability can be
XX monitored. The expression of genes can be used to study how FF cells
XX adapt to changes in culture conditions, environmental stress, spore
XX morphogenesis, recombination, metabolic or catabolic pathway
XX engineering. Using ESTs provides several advantages over genomic or
XX random cDNA clones including elimination of redundancy as one spot on an
XX array equals one gene or open reading frame, and organisation of the
XX microarrays based on function of the gene products to facilitate
XX analysis of the results. AAF07478 to AAF11247 represents ESTs from
XX *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*

Sequence 1611 BP; 326 A; 484 C; 422 G; 379 T; 0 other;

```

search completed: No
Job time : 466 secs

```

Search completed: November 21, 2003, 14:26:03

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2003, 20:31:12 ; Search time 122 Seconds
(without alignments)
6034.648 Million cell updates/sec

Title: US-09-712-338-1
Perfect score: 1668
Sequence: 1 atgcgtggctacgaattctt.....ccagtgtgtgtagcatag 1668

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1662	99.6	1662	3	US-08-943-714-1
2	86.4	5.2	2503	1	US-09-640-305-3
3	86.4	5.2	2503	1	US-08-360-673-3
4	84	5.0	2002	1	US-08-309-341-3
5	84	5.0	2002	1	US-08-608-267-3
6	84	5.0	2002	1	US-08-608-452-3
7	84	5.0	2002	1	US-08-608-224-3
8	84	5.0	2002	2	US-08-967-149-3
9	76	4.6	2068	1	US-08-309-341-1
10	76	4.6	2068	1	US-08-608-267-1
11	76	4.6	2068	1	US-08-608-452-1
12	76	4.6	2068	1	US-08-608-224-1
13	76	4.6	2068	2	US-08-967-149-1
14	66.2	4.0	2632	2	US-08-899-324-32
15	66.2	4.0	2632	3	US-08-329-892B-32
16	56.4	3.4	1814	4	US-09-702-705-319
17	56.4	3.4	1814	4	US-09-736-457-319
18	47.6	2.9	390	3	US-09-197-649-7
19	47.6	2.9	1551	2	US-08-828-488-4
20	47.6	2.9	1551	4	US-09-299-689A-4
21	47.2	2.8	1670	2	US-08-828-488-2
22	47.2	2.8	1670	4	US-09-299-689A-2
23	39.8	2.4	357	4	US-09-280-116-131
24	38.6	2.3	447	4	US-09-252-991A-11967
25	38.6	2.3	630	4	US-09-252-991A-6154
26	38.6	2.3	774	4	US-09-252-991A-11893
27	38.6	2.3	777	4	US-09-252-991A-8222

28	38.6	2.3	1083	4	US-09-252-991A-11930
29	38.6	2.3	1239	4	US-09-252-991A-8276
30	38.6	2.3	1434	4	US-09-252-991A-8046
31	38.2	2.3	918	3	US-09-105-390-35
32	38.2	2.3	1008	3	US-09-105-390-51
33	38.2	2.3	1949	3	US-09-105-390-2
34	37.2	2.2	870	4	US-09-252-991A-10938
35	37.2	2.2	1785	4	US-09-252-991A-11178
36	37.2	2.2	2091	4	US-09-252-991A-11076
37	36.8	2.2	7012	4	US-09-221-017B-173
38	36.4	2.2	1269	4	US-09-252-991A-8698
39	36.4	2.2	1443	4	US-09-252-991A-3554
40	36.4	2.2	1830	4	US-09-252-991A-3563
41	36.4	2.2	1926	4	US-09-252-991A-3531
42	36.4	2.2	1944	4	US-09-252-991A-9081
43	36.4	2.2	3494	3	US-09-139-802-200
44	36.4	2.2	3494	4	US-09-659-786-200
45	36.4	2.2	7218	1	US-08-232-463-14

ALIGNMENTS

RESULT 1

US-08-943-714-1

Sequence 1, Application US/08943714

Patent No. 6187578 **213101**

GENERAL INFORMATION:

APPLICANT: Blinkovsky, Alexander

APPLICANT: Berk, Randy

APPLICANT: Rev. Michael

APPLICANT: Goltz, Elizabeth

APPLICANT: Klotz, Alan

APPLICANT: Mathisen, Thomas Erik

APPLICANT: Dammann, Claus

TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESSES: Encoding Same

ADDRESSEE: No. 61875780 No. 6187578disk of No. 6187578th America, Inc.

STREET: 405 Lexington Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/943,714

FILING DATE: 03-OCT-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4990.200-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1662 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 1..1662

OTHER INFORMATION:

US-08-943-714-1									
Query Match 99.6%; Score 1662; DB 3; Length 1662;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1662; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	ATGCGTGCTACGAAATTTCTCTCAGTGTACCTTTGGTTGAGCCAGTCTGGCCCTTCCA	60						
DB	1	ATGCGTGCTACGAAATTTCTCTCAGTGTACCTTTGGTTGAGCCAGTCTGGCCCTTCCA	60						
QY	61	GGAAGTACACGGCGCTCGGTGAGACAGCTACCCAGAACCCACCGGGGTCAAG	120						
DB	61	GGAAGTACACGGCGCTCGGTGAGACAGCTACCCAGAACCCACCGGGGTCAAG	120						
QY	121	ACTCTTACAAACCGCAAAATGCTACCAATCCGGTACAAAGGAACCGGGGCGAGGGCGTC	180						
DB	121	ACTCTTACAAACCGCAAAATGCTACCAATCCGGTACAAAGGAACCGGGGCGAGGGCGTC	180						
QY	181	TGCGAGACTACCCCGGGGTCAAAATCTCTACTCTGGATATGTCGACACCTCTCCCGAGTCC	240						
DB	181	TGCGAGACTACCCCGGGGTCAAAATCTCTACTCTGGATATGTCGACACCTCTCCCGAGTCC	240						
QY	241	CATACCTTCTCTGTTCTTCGAGCCAGACATACCCAGAACTGCACCTATACATTC	300						
DB	241	CATACCTTCTCTGTTCTTCGAGCCAGACATACCCAGAACTGCACCTATACATTC	300						
QY	301	TGGTTGAATGGTGGCCCTGGAAGCGAATCTTTGATCGGTCTCTTCGAAAGATTGGGCCCT	360						
DB	301	TGGTTGAATGGTGGCCCTGGAAGCGAATCTTTGATCGGTCTCTTCGAAAGATTGGGCCCT	360						
QY	361	TGCGATGTCAATTCGACTTTTGGATGACTATACACCTCTCTCGTGGAACTGCTCC	420						
DB	361	TGCGATGTCAATTCGACTTTTGGATGACTATACACCTCTCTCGTGGAACTGCTCC	420						
QY	421	AATTTACTATCTCTGCTCCAGCAATGGGAGTCGGCTTTTCATATAGTGATACGTTGAT	480						
DB	421	AATTTACTATCTCTGCTCCAGCAATGGGAGTCGGCTTTTCATATAGTGATACGTTGAT	480						
QY	481	GGGTCCATTAACCTCTGAACTGGGGTCTGCGAAATTCGAGCTTTTCAGGAGTTCAGGGC	540						
DB	481	GGGTCCATTAACCTCTGAACTGGGGTCTGCGAAATTCGAGCTTTTCAGGAGTTCAGGGC	540						
QY	541	CGGTACCCAACTATGATGCCATCTGATGCTATACCAATCTTCGCGAGAGGCGCT	600						
DB	541	CGGTACCCAACTATGATGCCATCTGATGCTATACCAATCTTCGCGAGAGGCGCT	600						
QY	601	TGGGAGATCTCTGCAAGGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGGTGCAGTCT	660						
DB	601	TGGGAGATCTCTGCAAGGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGGTGCAGTCT	660						
QY	661	AAGGACTTCAGTCTATGCGAGAGCTATGAGGGGCACTATGGTCTCTGCAATCTTCAAT	720						
DB	661	AAGGACTTCAGTCTATGCGAGAGCTATGAGGGGCACTATGGTCTCTGCAATCTTCAAT	720						
QY	721	CATTTTACAGAGATGAGAAATGCGAAATGCGTATGTTTAAATGGTGTTCAGCTTAAT	780						
DB	721	CATTTTACAGAGATGAGAAATGCGAAATGCGTATGTTTAAATGGTGTTCAGCTTAAT	780						
QY	781	TTCAACTCTCTGGAAATTTATTAACGGCATCATCGAGGCGATCCAGGCGCCCTTACTAC	840						
DB	781	TTCAACTCTCTGGAAATTTATTAACGGCATCATCGAGGCGATCCAGGCGCCCTTACTAC	840						
QY	841	CCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGGTCAAC	900						
DB	841	CCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGGTCAAC	900						
QY	901	TACATGAAGTTTGCACCAAAATGCCAAATGGTTGCCAGGATTTGATTTCCACCTGCAA	960						
DB	901	TACATGAAGTTTGCACCAAAATGCCAAATGGTTGCCAGGATTTGATTTCCACCTGCAA	960						
QY	961	CAGACAAACCGCACCGGATAGCTGACTACGCCCTCTGCGCCGAGGACCAACATGTGC	1020						
DB	961	CAGACAAACCGCACCGGATAGCTGACTACGCCCTCTGCGCCGAGGACCAACATGTGC	1020						

RESULT 2
US-09-640-305-3
; Sequence 3, Application US/09640305
; Patent No. RE37447
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; Fournier, Alain
; Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3c43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/640,305
; FILING DATE: 16-Aug-2000
; PRIOR APPLICATION DATA:

TITLE OF INVENTION: GENE ENCODING CAREE

```

: 1. GENERAL INFORMATION:
:
: 2. APPLICANT: Yaver, Debbie Sue
:
: 3. APPLICANT: Thompson, Sheryl Ann
:
: 4. TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER

```

```

: 1. GENERAL INFORMATION:
:
: 2. APPLICANT: Yaver, Debbie Sue
:
: 3. APPLICANT: Thompson, Sheryl Ann
:
: 4. TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER

```

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSSEE: No. 55941190 No. 5594119disk of No. 5594119th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/309,341
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: intron
LOCATION: 349..411
FEATURE:
NAME/KEY: CDS
LOCATION: join (348..412)
US-08-309-341-3

Query Match 5.0%; Score 84; DB 1; Length 2002;
Best Local Similarity 55.5%; Pred. No. 2.8e-15;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 223 GACACCTCTCCCGAGTCCCATACCTTCTTCTGTTCTTTCGAGCCAGACATACCCAGAA 282
Db 617 GACACGAGAACGACAGCATCTTCTACTGTTCTTCGAGTCTCGAATGACCCGAG 676
QY 283 ACTGCACCTATCAATTTGTTGAATGTTGGCCCTGGAAGCGATTCTTTGATCGGTCTC 342
Db 677 AATGACCTTGTCTTCTGTTGTTGACGTTGGCCCTGGATGCTTCTCCCTCACCGTCTT 736
QY 343 TTCGAGAGTTGGGCCCTTCCCATGTCATTTGACATTTGATGACTACATCAACCTTAC 402
Db 737 TTCATGAGCTCGGCCCTAGCAGCATCAACAAGAGATCAGCGGCTCTACACGACTAC 796
QY 403 TCCTGGAACGAGTCTCCATTTACTATTCTCTCCAGCCATTGGGAGTCGGCTTTTCA 462
Db 797 GCTTGAACCTCCACGCTCGGTGATCTTCTTGAACGCTGTCAACGTCGGTTACTCT 856
QY 463 TATAGTATACGTTGATGGTTCATTAACCTTGAATGATGCTGGGTCTGCGA 512
Db 857 TACAGCAACTCTCTGTCTGACGACACCGTTGCTGTGGCAAGGACGCTA 906

RESULT 5
US-08-608-267-3
; Sequence 3, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann

TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/608,267
FILING DATE: 28-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,341
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: intron
LOCATION: 349..411
FEATURE:
NAME/KEY: CDS
LOCATION: join (348..412)
US-08-608-267-3

Query Match 5.0%; Score 84; DB 1; Length 2002;
Best Local Similarity 55.5%; Pred. No. 2.8e-15;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 223 GACACCTCTCCCGAGTCCCATACCTTCTTCTGTTCTTTCGAGCCAGACATACCCAGAA 282
Db 617 GACACGAGAACGACAGCATCTTCTACTGTTCTTCGAGTCTCGAATGACCCGAG 676
QY 283 ACTGCACCTATCAATTTGTTGAATGTTGGCCCTGGAAGCGATTCTTTGATCGGTCTC 342
Db 677 AATGACCTTGTCTTCTGTTGTTGACGTTGGCCCTGGATGCTTCTCCCTCACCGTCTT 736
QY 343 TTCGAGAGTTGGGCCCTTCCCATGTCATTTGACATTTGATGACTACATCAACCTTAC 402
Db 737 TTCATGAGCTCGGCCCTAGCAGCATCAACAAGAGATCAGCGGCTCTACACGACTAC 796
QY 403 TCCTGGAACGAGTCTCCATTTACTATTCTCTCCAGCCATTGGGAGTCGGCTTTTCA 462
Db 797 GCTTGAACCTCCACGCTCGGTGATCTTCTTGAACGCTGTCAACGTCGGTTACTCT 856
QY 463 TATAGTATACGTTGATGGTTCATTAACCTTGAATGATGCTGGGTCTGCGA 512
Db 857 TACAGCAACTCTCTGTCTGACGACACCGTTGCTGTGGCAAGGACGCTA 906

RESULT 6
US-08-608-452-3
; Sequence 3, Application US/08608452

Patent No. 5693510
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Thompson, Sheryl Ann
TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 56935100 No. 56935100 disk of No. 5693510th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/608,452
FILING DATE: 28-FEB-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: intron
LOCATION: 349..411
NAME/KEY: CDS
LOCATION: join (348..412)
US-08-608-452-3

Query Match 5.0%; Score 84; DB 1; Length 2002;
Best Local Similarity 55.5%; Pred. No. 2.8e-15;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 223 GACACCTCTCCGAGTCCCATACCTTCTTCTGTTCTTGAAGCCGACATACCCAGAA 282
Db 617 GACACGAGAACGACAGCATCTGTTCTTCTGAGTCTCGCAATGACCCGAG 676
QY 283 ACTGACCTATCATTTGTTGATGTTGGCCCTGGAAGCATTTTGTGCGTCTC 342
Db 677 AATGACCTGTTGTTCTGTTGATGAGTGGCCCTGGATGCTCTTCCCTCACCGTCTT 736
QY 343 TTCGAAAGTTGGGCCCTTGCCATGTCATTCGACTTTTTCATGACTACATCAACCTCTAC 402
Db 737 TTCATGAGCTCGGCCCTTAGCAGCATCAACAAGATCCAGCGGTCTACAGACTAC 796
QY 403 TCGTGAACGAGTCTCCAAATTTACTTCTTCCAGCCATTGGGAGTCGGCTTTTCA 462
Db 797 GCTTGAACCTCAACGCGTCCGTCATCTTCTTGAACGAGCTGTCAACGTCGGTTACTCT 856
QY 463 TATAGTGATCGGTTGATGGTCCATTAACCTGTAACTGGGTCGTCGA 512
Db 857 TACAGCAACTCTGCTGTACGACACCGTTGCTGCTGGCAAGACGCTCA 906

RESULT 7
US-08-608-224-3
Sequence 3, Application US/08608224
Patent No. 5705376
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Thompson, Sheryl Ann
TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 57053760 No. 57053760 disk of No. 5705376th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/608,224
FILING DATE: 28-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: intron
LOCATION: 349..411
NAME/KEY: CDS
LOCATION: join (348..412)
US-08-608-224-3

Query Match 5.0%; Score 84; DB 1; Length 2002;
Best Local Similarity 55.5%; Pred. No. 2.8e-15;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 223 GACACCTCTCCGAGTCCCATACCTTCTTCTGTTCTTGAAGCCGACATACCCAGAA 282
Db 617 GACACGAGAACGACAGCATCTGTTCTTCTGAGTCTCGCAATGACCCGAG 676
QY 283 ACTGACCTATCATTTGTTGATGTTGGCCCTGGAAGCATTTTGTGCGTCTC 342
Db 677 AATGACCTGTTGTTCTGTTGATGAGTGGCCCTGGATGCTCTTCCCTCACCGTCTT 736
QY 343 TTCGAAAGTTGGGCCCTTGCCATGTCATTCGACTTTTTCATGACTACATCAACCTCTAC 402
Db 737 TTCATGAGCTCGGCCCTTAGCAGCATCAACAAGATCCAGCGGTCTACAGACTAC 796
QY 403 TCGTGAACGAGTCTCCAAATTTACTTCTTCCAGCCATTGGGAGTCGGCTTTTCA 462
Db 797 GCTTGAACCTCAACGCGTCCGTCATCTTCTTGAACGAGCTGTCAACGTCGGTTACTCT 856

Qy 463 TATAGTGTACGGTTGATGGTGCCATTAACTCGGTCTGA 512
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dβ 857 TACAGCAACTCTGCTGCAGCGCACCGTTGCTGCTGGGAAGGACGTC TA 906
|||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8

US-08-967-149-3
; Sequence 3, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 59393050 No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowrey, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (348..412)
; US-08-967-149-3

	Query Match	5.0%;	Score 84;	DB 2;	Length 2002;
	Best Local Similarity	55.5%;	Mismat. No. 2.8e-15;		
	Matches 161;	Conservative 0;	Mismatches 129;	Indels 0;	Gaps 0
QY	223	GACACCTCCCGAGTCCCATACCTTCTCTGGTTCTTCGAAGCCAGACATATACCCGAA	282		
Db	617	GACACGAGAACGACCAAGCATCTGTCTACTGGTCTTCGAGTCTCGCAATGACCCGAG	676		
QY	283	ACTGCACCTATCACATTGTGGTTGAATGCTGGCCCTGGAAGCGATTCTTTGATCGGTCTC	342		
Db	677	AATGACCTGTGTCTGTGGTGACGCGTGGCCCTGGATGTCTTCCCTTACCGGTCTT	736		
OY	343	TTCGAAGAGCTTGGGCCCTTGCCATGTCAACTTTGATGACTACATCAACCCCTCAC	402		

Db	737	TTTCATGAGGCTCGGCCCTAGCAGCATCAACAAGAGATCCAGCCGGTCTACAAACGACTAC	796
Qy	403	TCGTGGAAACGAGGCTCCAAATTTACTATTCTGTGCCAGCCATTGGGAGTCGGCTTTTCA	462
Db	797	GCTTGGAACTCCAGGCGTCGTGATCTTCCTTGACCAGCCTGTCAACGTGCGTTACTCT	856
Qy	463	TATAGTGATACGGTTGATGGGTCCATTAAACCTGTAACCTGGGGTCGTGA	512
Db	857	TACAGCAACTCTGCTGTGACAGCACACCGTTGCTGCTGGCAGAGAGCTCTA	906

RESULT 9

US-08-309-341-1
; Sequence 1, Application US/08309341
; Patent No. 5594119
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5594119o No. 5594119disk of No. 5594119th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,341
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (571..633)
US-08-309-341-1

	Query Match	4.6%	Score 76	DB 1	Length 2068	
	Best Local Similarity	55.8%	Pred. No. 7.6e-13			
	Matches 145	Conservative 0	Mismatches 115	Indels 0	Gaps 0	
QY	253	TGGTCTTCGAAGCCGACATACCCAGAACTGCACCTATCATCTGGTGGTGAATGGT	312			
DB	633	TGGTCTTCGAGTCTCGCAATGACCCGAGAATATCCCGTTGTCTGTGGCTCAACGGT	692			
QY	313	GGCCCTGGAAGCGATTCTTTCATCGGTCTCTTTCGAAGAGTTGGGCCCTTGGCCATGTCAAT	372			
DB	693	GGCCCTGAGTCTCTTCCCTCACCGGTCTCTTCATGGAGCTTGGCCCTAGCAGCATCAAC	752			

QY 373 TCGACTTTGAGTACATCAACCCCTCACTCGTGGAGGAGTCTCCAAATTTACTATTTC 432
|||
Db 753 AAGAAGATCCAGCGGTCTTCAATGACTACGCTTGGAACTCCACGGGTCGGTGAATCTTC 812
|||
QY 433 CTGTCCAGCATTGGGAGTGGCTTTTCATATAGTGATAGGTTGATGGGTCCATTAAAC 492
|||
Db 813 CTTGACCAAGCTGTCAATGTGGTTACTCTCTACAGTAACCTGTGTCAGGACACGGTC 872
|||
QY 493 CTTGTAACCTGGGTCGTCGA 512
|||
Db 873 GCTGCTGGCAAGGACGTCCTA 892
|||
RESULT 10
US-08-608-267-1
; Sequence 1, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56886630 No. 56886630 disk of No. 56886630th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/08/608,267
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (571..633)
US-08-608-267-1

Query Match 4.6%; Score 76; DB 1; Length 2068;
Best Local Similarity 55.8%; Pred. No. 7.6e-13;
Matches 145; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 253 TGGTTCTTCGAGCCAGACATACCCAGAACTGCACCTATCAATGTGGTTGAATGGT 312
|||
Db 633 TGGTTCTTCGAGTCTCGCAATACCCGAGAAATGATCCCGTTGTTCTGTGGCTGAACGCT 692
|||

Query Match 4.6%; Score 76; DB 1; Length 2068;
Best Local Similarity 55.8%; Pred. No. 7.6e-13;
Matches 145; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 313 GGCCCTCGAAGCATTCTTTGATCGGTCTCTTCGAAAGTTGGCCCTTGCCATGTCAAT 372
|||
Db 693 GGCCCTCGGTGCTCTTCCCTCAACCGGTCTCTTCATGAGCTTGGCCCTAGCAGATCAAC 752
|||
QY 373 TCGACTTTGATGACTACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTACTATTTC 432
|||
Db 753 AAGAAGATCCAGCGGTCTTCAATGACTACGCTTGGAACTCCACGGGTCGGTGAATCTTC 812
|||
QY 433 CTGTCCAGCATTGGGAGTGGCTTTTCATATAGTGATAGGTTGATGGGTCCATTAAAC 492
|||
Db 813 CTTGACCAAGCTGTCAATGTGGTTACTCTCTACAGTAACCTGTGTCAGGACACGGTC 872
|||
QY 493 CTTGTAACCTGGGTCGTCGA 512
|||
Db 873 GCTGCTGGCAAGGACGTCCTA 892
|||

RESULT 11

US-08-608-452-1
; Sequence 1, Application US/08608452
; Patent No. 5693510
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56935100 No. 56935100 disk of No. 56935100th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/08/608,452
; APPLICATION NUMBER: US/08/608,452
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (571..633)
US-08-608-452-1

Qy	253	TGTTCTTCGAAGCAGACATAACCCAGAACTGCACCTATCACTGTGTGGTTGAATCGT	312
Db	633	TGTTCTTCGAGTCTCGCAATGACCCGAGAAATGATCCGTTGTTTCTGTGGCTGAACGCT	692
Qy	313	GGCCCTCGAAGCGGATCTTCATCGGTCTCTTCGAAGAGTTTGGGCCCTTGCCATCTCAAT	372
Db	693	GGCCCTGGTGCTCTTCCTCACCGGTCTCTTCATGGAGCTTGGCCCTTAGCAGCATCAAC	752
Qy	373	TCGACTTTTGTGACTACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTATATTC	432
Db	753	AAGAAGATCCAGCCGGTCTACAAATGACTACGCTTGGAACTCCAAACGCGTCCGTGATCTTC	812
Qy	433	CTGTCCAGACCATTTGGGAGTCGGCTTTTCATATAGTGATACGGTTGATCGGTCCTATTAAC	492
Db	813	CTTGACACGCTGTCAATGTTCGGTTACTCTTACAGTAACTCTGCTGTGACGACACGGTTC	872
Qy	493	CCTGTAACCTGGGGTGTGCGA	512
Db	873	GCTGCTGGCAAGGAGCTCTA	892

RESULT 12
 US-08-608-224-1
 ; Sequence 1, Application US/0808224
 ; Patent No. 5705376
 ; GENERAL INFORMATION:
 ; APPLICANT: Yaver, Debbie Sue
 ; APPLICANT: Thompson, Sheryl Ann
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5705376o No. 5705376disk of No. 5705376th America, Inc.
 ; STREET: 405 Lexington Avenue, Suite 6400
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/608,224
 ; FILING DATE: 28-FEB-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/309,341
 ; FILING DATE: 20-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lowney, Karen A.
 ; REGISTRATION NUMBER: 31,274
 ; REFERENCE/DOCKET NUMBER: 4247.000-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 867 0123
 ; TELEFAX: 212 867 0298
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2068 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Aspergillus niger
 ; FEATURE:
 ; NAME/KEY: intron
 ; LOCATION: 572...632
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: join (571..633)
 ; US-08-608-224-1

Query Match	4.68;	Score 76;	DB 1;	Length 2068;
Best Local Similarity	55.88;	Pred. No. 7.6e-13;	Indels	0;
Matches 145;	Conservative	0;	Mismatches 115;	Gaps 0
Qy	253	TGGTTCTTGGAAAGCCAGACATAACCCAGAAATCGCACCTATCAATGTGTGGTTGGAATGGT	312	
Db	633	TGGTTCTTCGAGTCTCGCAATGACCCCGAGAAATGATCCCGTGTTCCTGTGGCTGCAACGGT	692	
Qy	313	GGCCCTGGAGAGCGATTCTTTGATCGGTCCTCTTCGAAGAGTTGGGCCCTTGGCCATGTCAT	372	
Db	693	GGCCCTGGGTGCTCTTCCCTACCGGTCCTTCATGGAGCTTGGCCCTAGCAGCATCAAC	752	
Qy	373	TCGACTTTTGATGACTACATCAACCCCTCACTCGTGGAACGAGGCTTCCAAATTTACTATTC	432	
Db	753	AAGAAGATCCAGCCGGTCTACATGACTACGCTTGGAACTCCAAACGCGTCGGTGATCTTC	812	
Qy	433	CTGTCCAGACCATTTGGGAGTCGGCTTTTCATATAGTAGTACGTTGATGGGTCCAAATAAC	492	
Db	813	CTTGACAGACCTGTCAATGTCGGTTACTCCTACAGTAACCTGCTGTCAAGGCACACGGTC	872	
Qy	493	CCTGTAACCTGGGGTCGTGA	512	
Db	873	GCTGCTGGCAAGGACGTCTA	892	

RESULT 13
 US-08-967-149-1
 ; Sequence 1, Application US/08967149
 ; Patent No. 5939305
 ; GENERAL INFORMATION:
 ; APPLICANT: Yaver, Debbie Sue
 ; APPLICANT: Thompson, Sheryl Ann
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
 ; TITLE OF INVENTION: ASPERGILLUS NIGER
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5939305 No. 5939305disk of No. 5939305th America, Inc.
 ; STREET: 405 Lexington Avenue, Suite 6400
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/967,149
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/608,452
 ; FILING DATE: 28-FEB-1996
 ; APPLICATION NUMBER: US 08/309,341
 ; FILING DATE: 20-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lowney, Karen A.
 ; REGISTRATION NUMBER: 31,274
 ; REFERENCE/DOCKET NUMBER: 4247.000-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 867 0123
 ; TELEFAX: 212 867 0298
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2068 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Aspergillus niger
 ; FEATURE:

```
/ NAME/KEY: intron
/ LOCATION: 572..632
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: join (571...633)
US-08-967-149-1

Query Match
Best Local Similarity 4.6%; Score 76; DB 2; Length 2068;
Matches 145; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 253 TGGTCTTCGAGCCAGACATACCCGAACTCACTATCATTGTGTTGAATGGT 312
Db |||||
633 TGGTCTTCGAGCTCGCAATGACCCGAGAATGATCCCGTTGTCTGTGCTGAACGGT 692
QY 313 GGCCTCGAAGCGATTCTTTGATCGGTCTCTTCAAGAGTTGGCCCTTGCATGTAAT 372
Db |||||
693 GGCCTCGGTGCTCTTCCCTCACCCTGCTCTTCAAGAGCTTGGCCCTAGCAGCATCAAC 752
QY 373 TCGACTTTTGATGACTACATCAACCTCACTCGTGGAAAGAGGTCTCCAAATTTACTATTC 432
Db |||||
753 AAGAAGATCCAGCGGCTCTACAATGACTACGCTTGGAACTCCACGCGTCCGTGATCTTC 812
QY 433 CTGTCCAGCATTTGGAGTCGGCTTTTCAATAGTATGATACGTTGATGGTCCATTAAC 492
Db |||||
813 CTTGACAGCGCTGTCAATGCTGCTTACTCTCTACAGTAACCTCTGCTGACGACACGGTC 872
QY 493 CCGTAACTGGGTCTGTCGA 512
Db |||||
873 GCTGCTGGCAAGGACGCTA 892

RESULT 14
US-08-899-324-32
; Sequence 32, Application US/08899324
; Patent No. 5945329
; GENERAL INFORMATION:
; APPLICANT: Breddam, Klaus
; APPLICANT: Keiland-Brandt, Morten
; APPLICANT: Mortensen, Uffe
; APPLICANT: Olesen, Kjeld
; APPLICANT: Stennicke, Henning
; APPLICANT: Wagner, Fred
; TITLE OF INVENTION: CUSTOMIZED PROTEASES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 5945329west Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,324
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/329,892
; FILING DATE: 27-OCT-1994
; APPLICATION NUMBER: 08/144,704
; FILING DATE: 28-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettleberger, Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 8648.44USC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
```

```
TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2632 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 696...2291
; OTHER INFORMATION:
US-08-899-324-32

Query Match
Best Local Similarity 4.0%; Score 66.2; DB 2; Length 2632;
Matches 140; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 241 CATACCTTCTTCGGTCTTCGAAAGCCAGACATACCCGAACTGCACCTATCACATTG 300
Db |||||
1113 CATTTCTCTTTGGACTTTTGAAGTAGAAAGATCCTGCAAGGATCCGGTCATCCTT 1172
QY 301 TGGTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTTCGAAGAGTTGGGCCCT 360
Db |||||
1173 TGGTGAACGGGGTCCAGGTCTTCTTCACTAACCGGGCTGTTCTTTGAATTAGGACCC 1232
QY 361 TGCCATGTCAATTCGACTTTTGATGACTACATCAACCTCAGTCTGGAACGAGGTCTCC 420
Db |||||
1233 TCATCCATTGGACCTGATTTGAACCCATCGGAACCCCTTACTCTTGAACAGCAATGCC 1292
QY 421 AATTACTATCTCTGCCAGCCATTGGGAGTGGGCTTTTTCATATAGTAGATCGGTGAT 480
Db |||||
1293 ACGTGTATCTCTTCACACGCTGTCACGTTGGGTTCGTATTCCGGGTCTCTCAGGT 1352
QY 481 GGTCCATTAAACCTGTAACTGG 503
Db |||||
1353 GTTCCAACTGTCGCGCTGG 1375

RESULT 15
US-08-329-892B-32
; Sequence 32, Application US/08329892B
; Patent No. 6187579
; GENERAL INFORMATION:
; APPLICANT: Breddam, Klaus
; APPLICANT: Keiland-Brandt, Morten
; APPLICANT: Mortensen, Uffe
; APPLICANT: Olesen, Kjeld
; APPLICANT: Stennicke, Henning
; APPLICANT: Wagner, Fred
; TITLE OF INVENTION: CUSTOMIZED PROTEASE
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6187579west Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/329,892B
; FILING DATE: 27-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 13:23:32 ; Search time 399 Seconds
(without alignments)
11284.869 Million cell updates/sec

Title: US-09-712-338-1
Perfect score: 1668
Sequence: 1 atgcgtggctacgaatttct.....ccagttgtgtatgcatag 1668

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq, 19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1662	99.6	1662	19 AAV28620	A. oryzae ATCC2038
2	25	1.5	1656	24 ABZ78288	A. niger serine ca
3	25	1.5	3150	24 ABZ78231	A. niger serine ca
4	21	1.3	1515	21 AAA70412	Synthetic HIV Gag
5	21	1.3	1539	24 AA167899	Synthetic HIV Gag
6	21	1.3	1539	24 AA167900	Nucleotide sequenc
7	21	1.3	2031	21 AAA70415	Synthetic HIV Gag/
8	21	1.3	2799	25 ACA03520	Synthetic DNA ecod

9	21	1.3	2799	25 ACA03583	Synthetic DNA ecod
10	21	1.3	4472	21 AAA70472	HIV bicistronic co
11	21	1.3	4608	21 AAA70473	HIV bicistronic co
12	21	1.3	4689	21 AAA70471	HIV bicistronic co
13	21	1.3	4766	21 AAA70470	HIV bicistronic co
14	21	1.3	4773	25 ACA03519	Synthetic DNA ecod
15	21	1.3	4773	25 ACA03585	Synthetic DNA ecod
16	21	1.3	5274	25 ACA03518	Synthetic DNA ecod
17	21	1.3	5283	25 ACA03584	Synthetic DNA ecod
18	19	1.1	989	24 ABK73951	Bacillus lichenifo
19	19	1.1	11838	23 ABL16320	Drosophila melanog
20	18	1.1	299	25 ABX29819	Human GDP-mannose
21	18	1.1	390	21 ABX67203	Pinus radiata alph
22	18	1.1	416	21 AAZ42530	Human 5' EST isola
23	18	1.1	441	21 AA667198	Pinus radiata alph
24	18	1.1	447	22 AA188892	Human polynucleoti
25	18	1.1	481	21 AA667199	Pinus radiata alph
26	18	1.1	602	25 ABX05179	Human novel polynu
27	18	1.1	614	21 AAZ42528	Human 5' EST isola
28	18	1.1	637	22 AA331039	Human diagnostic a
29	18	1.1	637	24 ABQ72572	Human MDDT encodin
30	18	1.1	660	21 AAC44631	Arabidopsis thalia
31	18	1.1	696	22 AA160959	Human polynucleoti
32	18	1.1	700	24 ABQ31888	Oligonucleotide fo
33	18	1.1	700	24 ABQ31889	Oligonucleotide fo
34	18	1.1	703	24 ABQ33578	Oligonucleotide fo
35	18	1.1	703	24 ABQ33579	Oligonucleotide fo
36	18	1.1	716	22 AA159173	Human polynucleoti
37	18	1.1	1486	23 ABL16753	Drosophila melanog
38	18	1.1	1572	23 ABL09011	A. niger serine ca
39	18	1.1	1581	24 ABZ78241	Drosophila melanog
40	18	1.1	1702	23 ABL16073	DNA encoding type
41	18	1.1	2592	18 AA162651	Human cDNA differe
42	18	1.1	2592	24 ABK84745	A. niger serine ca
43	18	1.1	2940	24 ABZ78184	Human DNA sequence
44	18	1.1	3128	24 AA594904	Aspergillus fumiga
45	18	1.1	3380	25 ABT17664	

ALIGNMENTS

RESULT 1
AAV28620
ID AAV28620 standard; DNA; 1662 BP.

XX AAV28620;

XX AC AAV28620;

XX DT 27-AUG-1998 (first entry)

XX DE A. oryzae ATCC20386 carboxypeptidase I DNA.

XX DE Carboxypeptidase I; flavour improving agent; hydrolasate; proteinaceous;

XX KW food industry; ss.

XX OS Aspergillus oryzae.

XX FH Key Location/Qualifiers

XX CDS 1..1662

XX FT /*tag= a

XX FT /product= carboxypeptidase I

XX FT /note= "partial coding sequence"

XX PN WO9814599-A1.

XX PD 09-APR-1998.

XX PF 03-OCT-1997; 97WO-US17977.

XX PR 27-NOV-1996; 96US-0757534.

XX PR 04-OCT-1996; 96US-0736880.

XX PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO-NORDISK AS.

XX Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;

PI Klotz A, Mathisen TE, Rey M;

XX WPI; 1998-240098/21.

DR P-ESDB; AAW56099.

XX Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous
PT substrates, useful for improving flavour of foods

XX Claim 2; Fig 3; 82pp; English.

XX This DNA sequence encodes carboxypeptidase I from Aspergillus oryzae.
CC This polypeptide has an optimal activity in the range of pH 3.0-7.5 at
CC 25 deg. C. Optimal activity in the range of 55-60 deg. C at pH 4, and a
CC residual activity of at least 65.5% after 30 minutes at pH 4.0 and
CC 60 deg. C. It also has the capacity to hydrolyse X from N-CBZ-Ala-X where
CC N-CBZ is N-carboxybenzyl and X is any amino acid. The carboxypeptidases
CC can be used for obtaining hydrolysates (which can be enriched in free
CC glutamic acid or peptide bound glutamic acid residues) from proteinaceous
CC substrates. The carboxypeptidases can be used in flavour-improving
CC compositions in the food industry. The products can also be used for the
CC production of polypeptides free of carboxypeptidase activity.

XX SQ Sequence 1662 BP; 396 A; 468 C; 398 G; 400 T; 0 other;

Query Match 99.6%; Score 1662; DB 19; Length 1662;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1662; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGTGGCTACGAATTTCTCTCAGTGTCTACCTTGGTGGAGCCAGTGGGCCCTTCCA 60

DB 1 ATGCGTGGCTACGAATTTCTCTCAGTGTCTACCTTGGTGGAGCCAGTGGGCCCTTCCA 60

QY 61 GGAAGTACACCGGCTCGGTAGAGACAGTACCCAGAACCCACCGGGTCAAG 120

DB 61 GGAAGTACACCGGCTCGGTAGAGACAGTACCCAGAACCCACCGGGTCAAG 120

QY 121 ACTCTTAAACCGCAAAACAATGTCAACATCCCGGTACAAAGAACCCGGGCGAGGGCGTC 180

DB 121 ACTCTTAAACCGCAAAACAATGTCAACATCCCGGTACAAAGAACCCGGGCGAGGGCGTC 180

QY 181 TGCAGAGTACACCGGGTGTCAAAATCTCTCTGATATGTGCAACCTCTCCGAGTCC 240

DB 181 TGCAGAGTACACCGGGTGTCAAAATCTCTCTGATATGTGCAACCTCTCCGAGTCC 240

QY 241 CATACCTTCTCTGTTCTTTCGAAGCCAGACATACCCAGAACCTGACCTATCATATG 300

DB 241 CATACCTTCTCTGTTCTTTCGAAGCCAGACATACCCAGAACCTGACCTATCATATG 300

QY 301 TGGTTGAATGGTGGCCCTCGAAGCGATTCTTTGATCGGTCTCTTCGAAGTGGGCCCT 360

DB 301 TGGTTGAATGGTGGCCCTCGAAGCGATTCTTTGATCGGTCTCTTCGAAGTGGGCCCT 360

QY 361 TGCATGTCAATTCGACTTTTGAATGACTATCAACCTCTCTCGTGGAACGAGTCTCC 420

DB 361 TGCATGTCAATTCGACTTTTGAATGACTATCAACCTCTCTCGTGGAACGAGTCTCC 420

QY 421 AATTTTACTTCTCTGTCGCCAGCATTTGGAGTGGGCTTTTATATAGTATCATACGTTGAT 480

DB 421 AATTTTACTTCTCTGTCGCCAGCATTTGGAGTGGGCTTTTATATAGTATCATACGTTGAT 480

QY 481 GGGTCCATTAACCTCTGTAATCGGGTGTCTGAAAATTCGAAATTCGAGCTTTCAGGGC 540

DB 481 GGGTCCATTAACCTCTGTAATCGGGTGTCTGAAAATTCGAAATTCGAGCTTTCAGGGC 540

QY 541 CGGTACCCAAACCATTTGATGGCATCTGATGATACCTACCAATCTTGGCCGACAGGCCGT 600

DB 541 CGGTACCCAAACCATTTGATGGCATCTGATGATACCTACCAATCTTGGCCGACAGGCCGT 600

QY 601 TGGGAGATCCTGCAAGGATTCCTTGTAGTGAGTACCTAGTCTGAGCTCTAGGGTGGAGTCT 660

RESULT 2
ABZ78288

DB 601 TGGAGATCCTGCAAGGATTCCTTAGTGGACTACTAGCTGGACTTAGGGTGCAGTCT 660

QY 661 AAGGACTTCACTCTATGAGAGGAGAGCTATGAGGGCACTATGFCCTGCAATCTTCAAT 720

DB 661 AAGGACTTCACTCTATGAGAGGAGAGCTATGAGGGCACTATGFCCTGCAATCTTCAAT 720

QY 721 CATTTTACGAGAGAGAGAGATTCGCAACGGTAGTGTAAATGTTTACGCTTAAT 780

DB 721 CATTTTACGAGAGAGAGAGATTCGCAACGGTAGTGTAAATGTTTACGCTTAAT 780

QY 781 TTCAACTCTCTGGCAATTTATTAACGGCATCATCAAGAGGAGATTCAGGCCCTTACTAC 840

DB 781 TTCAACTCTCTGGCAATTTATTAACGGCATCATCAAGAGGAGATTCAGGCCCTTACTAC 840

QY 841 CCTCAATTCGTGTGAACAATACCTACGGTATCAAGGTGTCAACGAGACCGTCTTCAAC 900

DB 841 CCTCAATTCGTGTGAACAATACCTACGGTATCAAGGTGTCAACGAGACCGTCTTCAAC 900

QY 901 TACATGAATTTGCCAACCAATGCGAATGGTTCGCGAGGATTTGATTTCCACCTGCAAA 960

DB 901 TACATGAATTTGCCAACCAATGCGAATGGTTCGCGAGGATTTGATTTCCACCTGCAAA 960

QY 961 CAGACAAACCGCACCGCATTAGCTGACTACGCCCTCTCGCGGAGGACCAACCAATGTGC 1020

DB 961 CAGACAAACCGCACCGCATTAGCTGACTACGCCCTCTCGCGGAGGACCAACCAATGTGC 1020

QY 1021 AGGACAATTTGAGGGGCCATACCTACGCCCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080

DB 1021 AGGACAATTTGAGGGGCCATACCTACGCCCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080

QY 1081 CATCCATATGATGACCGGACTCCGCCAAGTTATTACAAATTTCTGSCAAAGGACTCT 1140

DB 1081 CATCCATATGATGACCGGACTCCGCCAAGTTATTACAAATTTCTGSCAAAGGACTCT 1140

QY 1141 GTCATGAGCGCTATCGGGTCAACATCAACTACACCCAGTCCAAATATGAGCTCTACTAC 1200

DB 1141 GTCATGAGCGCTATCGGGTCAACATCAACTACACCCAGTCCAAATATGAGCTCTACTAC 1200

QY 1201 GCTTTCAGCAAAACGAGCGACTTTGTCTGGGCCAACTTCTGAGAGAGCTCGAGAGATC 1260

DB 1201 GCTTTCAGCAAAACGAGCGACTTTGTCTGGGCCAACTTCTGAGAGAGCTCGAGAGATC 1260

QY 1261 CTTGCTCTCCCGTGGCTCTCCCTCATCTATGCGAGCGGATTTACATCTGCAACTGG 1320

DB 1261 CTTGCTCTCCCGTGGCTCTCCCTCATCTATGCGAGCGGATTTACATCTGCAACTGG 1320

QY 1321 TTCCGCGGTCAAGGCCGTTTCCCTCTGCTCGAACTACTCCCAAGCGCCAGTTCGGAAGC 1380

DB 1321 TTCCGCGGTCAAGGCCGTTTCCCTCTGCTCGAACTACTCCCAAGCGCCAGTTCGGAAGC 1380

QY 1381 GCAGGGTACAGCCCTGAAAGTCAACGGGTGAGTATGGGAACTCGCGAGTATGGT 1440

DB 1381 GCAGGGTACAGCCCTGAAAGTCAACGGGTGAGTATGGGAACTCGCGAGTATGGT 1440

QY 1441 AATTTTCTCTTCACTCGCGTCTATGAGGAGGAGGATGAGTCCCATCTACAGGCCATC 1500

DB 1441 AATTTTCTCTTCACTCGCGTCTATGAGGAGGAGGATGAGTCCCATCTACAGGCCATC 1500

QY 1501 GCCTCCCTGCAATTTTAAACCGGACTATCTTCGGTGGGATATCCGAGAGGCCAGAG 1560

DB 1501 GCCTCCCTGCAATTTTAAACCGGACTATCTTCGGTGGGATATCCGAGAGGCCAGAG 1560

QY 1561 AAGATCTGGCCAGGTACAAAGCGAATGGAGCGGTACAGCTACAGTACAGTACAGTCC 1620

DB 1561 AAGATCTGGCCAGGTACAAAGCGAATGGAGCGGTACAGCTACAGTACAGTACAGTCC 1620

QY 1621 GTGCCGCTGCCTACCGGTACCAAGCATGTCCAGTGTGGTATG 1662

DB 1621 GTGCCGCTGCCTACCGGTACCAAGCATGTCCAGTGTGGTATG 1662

ID ABZ78288 standard; cDNA; 1656 BP.
 AC ABZ78288;
 XX
 DT 24-APR-2003 (first entry)
 XX
 DE A. niger serine carboxypeptidase cDNA #3.
 XX
 XX Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ss.
 XX
 OS Aspergillus niger.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1656
 FT /tag= a
 FT /EC_number= "3.4.16.6"
 XX
 XX WO200268623-A2.
 XX
 XX 06-SEP-2002.
 XX
 XX 22-FEB-2002; 2002WO-EP01984.
 XX
 PR 23-FEB-2001; 2001EP-0200657.
 PR 23-FEB-2001; 2001EP-0200658.
 PR 23-FEB-2001; 2001EP-0200660.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 26-FEB-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0000075.
 PR 28-MAR-2001; 2001EP-0000078.
 PR 28-MAR-2001; 2001EP-0000080.
 PR 28-MAR-2001; 2001EP-0000087.
 PR 28-MAR-2001; 2001EP-0000088.
 PR 21-MAY-2001; 2001EP-0000156.
 PR 21-MAY-2001; 2001EP-0000159.
 PR 21-MAY-2001; 2001EP-0000160.
 PR 21-MAY-2001; 2001EP-0000162.
 PR 21-MAY-2001; 2001EP-0000165.
 PR 21-MAY-2001; 2001EP-0000166.
 PR 21-MAY-2001; 2001EP-0000168.
 PR 21-JUN-2001; 2001EP-0000240.
 PR 21-JUN-2001; 2001EP-0000242.
 PR 21-JUN-2001; 2001EP-0000244.
 PR 21-JUN-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 12-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000323.
 PR 30-JUL-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 22-OCT-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0004464.
 PR 21-DEC-2001; 2001EP-0005117.
 PR (STAM) DSM NV.
 PR Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Stiebler J, Albang R;
 XX
 DR WPI; 2002-723203/78.
 DR P-PSDB; ABR38964.
 XX
 PT Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications
 XX
 PS Claim 1; Page 252-253; 394pp; English.
 XX
 CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 CC A. niger proteases of the invention.
 XX
 SQ Sequence 1656 BP; 411 A; 426 C; 409 G; 410 T; 0 other;
 Query Match 1..5%; Score 25; DB 24; Length 1656;
 Best Local Similarity 100.0%; Pred. No. 0.032;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 305 TGAATGGTGGCCCTGGAGCGATTC 329
 Db 305 TGAATGGTGGCCCTGGAGCGATTC 329
 RESULT 3
 ABZ78231
 ID ABZ78231 standard; DNA; 3150 BP.
 XX
 AC ABZ78231;
 XX
 DT 24-APR-2003 (first entry)
 XX
 DE A. niger serine carboxypeptidase gene #3.
 XX
 KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ds.
 XX
 OS Aspergillus niger.
 XX
 XX WO200268623-A2.
 XX
 PD 06-SEP-2002.
 XX
 PF 22-FEB-2002; 2002WO-EP01984.
 XX
 PR 23-FEB-2001; 2001EP-0200657.
 PR 23-FEB-2001; 2001EP-0200658.
 PR 23-FEB-2001; 2001EP-0200660.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 26-FEB-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0000075.
 PR 28-MAR-2001; 2001EP-0000078.
 PR 28-MAR-2001; 2001EP-0000080.
 PR 28-MAR-2001; 2001EP-0000087.
 PR 28-MAR-2001; 2001EP-0000088.
 PR 21-MAY-2001; 2001EP-0000156.
 PR 21-MAY-2001; 2001EP-0000159.
 PR 21-MAY-2001; 2001EP-0000160.
 PR 21-MAY-2001; 2001EP-0000162.
 PR 21-MAY-2001; 2001EP-0000165.
 PR 21-MAY-2001; 2001EP-0000166.
 PR 21-MAY-2001; 2001EP-0000168.
 PR 21-JUN-2001; 2001EP-0000240.
 PR 21-JUN-2001; 2001EP-0000242.
 PR 21-JUN-2001; 2001EP-0000244.
 PR 21-JUN-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 12-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000323.
 PR 30-JUL-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 22-OCT-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0004464.
 PR 21-DEC-2001; 2001EP-0005117.
 PR (STAM) DSM NV.
 PR Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Stiebler J, Albang R;
 XX
 DR WPI; 2002-723203/78.
 DR P-PSDB; ABR38964.
 XX
 PT Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications
 XX
 PS Claim 1; Page 252-253; 394pp; English.
 XX
 CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 CC A. niger proteases of the invention.
 XX
 SQ Sequence 1656 BP; 411 A; 426 C; 409 G; 410 T; 0 other;
 Query Match 1..5%; Score 25; DB 24; Length 1656;
 Best Local Similarity 100.0%; Pred. No. 0.032;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 305 TGAATGGTGGCCCTGGAGCGATTC 329
 Db 305 TGAATGGTGGCCCTGGAGCGATTC 329

PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-JUN-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.
PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 30-JUL-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.
PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000344.
PR 09-AUG-2001; 2001EP-0000357.
PR 16-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.
PR 22-OCT-2001; 2001EP-0000556.
PR 22-OCT-2001; 2001EP-0000557.
PR 22-OCT-2001; 2001EP-0000558.
PR 15-NOV-2001; 2001EP-0004464.
PR 21-DEC-2001; 2001EP-0005117.
XX
PA (STAM) DSM NV.

XX
XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albarg R;
XX
XX WPI; 2002-723203/78.
DR P-PSDB; ABR38864.

XX
XX Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications -
XX

PS Claim 1; Page 187-189; 394pp; English.

XX
XX The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing, in
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the
CC A. niger proteases of the invention.
XX

SQ Sequence 3150 BP; 801 A; 778 C; 738 G; 833 T; 0 other;

Query Match 1.5%; Score 25; DB 24; Length 3150;
Best Local Similarity 100.0%; Pred.No. 0.032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 TGAATGGTGGCCCTGGAGCGATTC 329
|||||
DB 879 TGAATGGTGGCCCTGGAGCGATTC 903

RESULT 4

AAA70412
ID AAA70412 standard; DNA; 1515 BP.

XX
AC AAA70412;

XX 28-NOV-2000 (first entry)

XX Synthetic HIV Gag expression cassette coding sequence Gag.ModSF2.

XX HIV-1; AIDS; Gag; vaccine; expression cassette; ss.

XX Human immunodeficiency virus type 1.
OS Synthetic.

XX WO2000039302-A2.

XX 06-JUL-2000.

XX 30-DEC-1999; 99WO-US31245.

XX 31-DEC-1998; 98US-0114495.

XX 01-DEC-1999; 99US-0168471.

XX (CHIR) CHIRON CORP.

XX Barnett S, Zur Megede J, Srivastava I, Lian Y, Hartog K, Liu H;
PI Greer C, Selby M, Walker C;

XX WPI; 2000-452400/39.

XX Expression cassettes encoding the human immunodeficiency virus (HIV)
PT Gag-containing polypeptide useful for vaccinating against HIV
PT infections and acquired immunodeficiency syndrome (AIDS) -

PS Claim 3; Fig 7; 391pp; English.

XX The present sequence is the coding sequence of a HIV Gag expression
CC cassette, Gag.ModSF2. The Gag protein of HIV is needed for the assembly
CC of virus-like particles. In addition, the Gag protein is involved in
CC many stages of the HIV life cycle, including assembly, virion maturation
CC after particle release and early post-entry steps in viral replication.
CC The expression cassette may be used for the recombinant expression of
CC HIV Gag-polypeptides which may then be used to vaccinate against HIV
CC infection and acquired immunodeficiency syndrome (AIDS).
XX

SQ Sequence 1515 BP; 329 A; 547 C; 480 G; 159 T; 0 other;

Query Match 1.3%; Score 21; DB 21; Length 1515;
Best Local Similarity 100.0%; Pred.No. 3.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGCGCCGAGCTACAAG 1581

DB 1318 AAGATCTGCGCCGAGCTACAAG 1338

RESULT 5

AAI67899

ID AAI67899 standard; DNA; 1539 BP.

XX AAI67899;

XX 13-MAR-2002 (first entry)

XX Nucleotide sequence of a synthetic HIV-1 gag coding region.

XX Gag; gagpol; pol; codon optimisation; virucide; immunostimulant;
KW vaccine; gene therapy; siV; HIV; ss.

XX Synthetic.

```

OS Human immunodeficiency virus type 1.
XX
PN EP1156112-A1.
XX
PD 21-NOV-2001.
XX
XX 18-MAY-2000; 2000EP-0110623.
XX
PF 18-MAY-2000; 2000EP-0110623.
XX
PR 18-MAY-2000; 2000EP-0110623.
XX
PA (GENE-) GENEART GMBH.
XX
XX
PI Wagner R, Graf M, Deml L, Bieler K;
XX
DR WPI; 2002-043171/06.
XX
XX
PT Synthetic gag and gagpol genes optimized for high level expression via
PT codon optimization, for efficient generation of vector particle useful
PT in gene delivery and for immunizing against diseases of retroviruses -
XX
PS Example 1; Page 26; 40pp; English.
XX
CC The invention relates to synthetic gag and gagpol genes optimized for
CC high level expression via codon optimization. The optimised genes can be
CC used for treatment and prophylaxis of diseases caused by retroviruses as
CC well as efficient generation of vector particles (as packaging construct
CC for gene delivery). The optimized genes may also be used in the
CC construction of GagPol based vaccines modified for improved safety. When
CC used as packaging constructs, the optimised genes exhibit safe and
CC efficient gene delivery and expression of Gag of GagPol genes; and
CC allows expression of high yields of native proteins in higher eukaryotes
CC or insect cells, and supporting the generation of transduction competent
CC particles for vaccination and gene therapy purposes, unlike the complex
CC regulatory mechanisms of late gag and gagpol expression found in the
CC prior art. The present sequence represents the nucleotide sequence of
CC a synthetic HIV-1 gag coding region.
XX
SQ Sequence 1539 BP; 366 A; 508 C; 502 G; 163 T; 0 other;
Query Match 1.3%; Score 21; DB 24; Length 1539;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1561 AAGATCTGCCCGACTACAAG 1581
Db |||||
1306 AAGATCTGCCCGACTACAAG 1326

RESULT 6
AAI67900
ID AAI67900 standard; DNA; 1539 BP.
XX
AC AAI67900;
XX
DT 13-MAR-2002 (first entry)
XX
DE Nucleotide sequence of a synthetic HIV-1 gag coding region.
XX
KW Gag; gagpol; pol; codon optimisation; virucide; immunostimulant;
KW vaccine; gene therapy; SIV; HIV; ss.
XX
XX Synthetic.
OS Human immunodeficiency virus type 1.
XX
PN EP1156112-A1.
XX
PD 21-NOV-2001.
XX
XX 18-MAY-2000; 2000EP-0110623.
XX
PF 18-MAY-2000; 2000EP-0110623.
XX
PR 18-MAY-2000; 2000EP-0110623.
XX
PA (GENE-) GENEART GMBH.
XX
XX
PI Wagner R, Graf M, Deml L, Bieler K;
XX
DR WPI; 2002-043171/06.
XX
XX
PT Synthetic gag and gagpol genes optimized for high level expression via
PT codon optimization, for efficient generation of vector particle useful
PT in gene delivery and for immunizing against diseases of retroviruses -
XX
PS Example 1; Page 26; 40pp; English.
XX
CC The invention relates to synthetic gag and gagpol genes optimized for
CC high level expression via codon optimization. The optimised genes can be
CC used for treatment and prophylaxis of diseases caused by retroviruses as
CC well as efficient generation of vector particles (as packaging construct
CC for gene delivery). The optimized genes may also be used in the
CC construction of GagPol based vaccines modified for improved safety. When
CC used as packaging constructs, the optimised genes exhibit safe and
CC efficient gene delivery and expression of Gag of GagPol genes; and
CC allows expression of high yields of native proteins in higher eukaryotes
CC or insect cells, and supporting the generation of transduction competent
CC particles for vaccination and gene therapy purposes, unlike the complex
CC regulatory mechanisms of late gag and gagpol expression found in the
CC prior art. The present sequence represents the nucleotide sequence of
CC a synthetic HIV-1 gag coding region.
XX
SQ Sequence 1539 BP; 366 A; 508 C; 502 G; 163 T; 0 other;
Query Match 1.3%; Score 21; DB 24; Length 1539;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1561 AAGATCTGCCCGACTACAAG 1581
Db |||||
1306 AAGATCTGCCCGACTACAAG 1326

RESULT 7
AAI70415
ID AAI70415 standard; DNA; 2031 BP.
XX
AC AAI70415;
XX
DT 28-NOV-2000 (first entry)
XX
DE Synthetic HIV Gag/HCV core fusion coding sequence.
XX
KW HIV-1; AIDS; Gag; vaccine; expression cassette; ss.
XX
OS Chimeric - Human immunodeficiency virus type 1.
OS Chimeric - Hepatitis C virus.
XX
PN WO200039302-A2.
XX
PD 06-JUL-2000.
XX
XX 30-DEC-1999; 99WO-US31245.
XX
PR 31-DEC-1998; 98US-0114495.
PR 01-DEC-1999; 99US-0168471.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Barnett S, Zur Megede J, Srivastava I, Lian Y, Hartog K, Liu H;
PI Greer C, Selby M, Walker C;
XX
XX WPI; 2000-452400/39.
XX
XX Expression cassettes encoding the human immunodeficiency virus (HIV)
XX Gag-containing polypeptide useful for vaccinating against HIV
XX infections and acquired immunodeficiency syndrome (AIDS) -
XX
PS Example 1; Pages 341-342; 391pp; English.

```

XX The present sequence is a HIV Gag/Hepatitis C virus (HCV) core fusion
 CC coding sequence. The Gag protein of HIV is needed for the assembly of
 CC virus-like particles. In addition, the Gag protein is involved in many
 CC stages of the HIV life cycle, including assembly, virion maturation after
 CC particle release and early post-entry steps in viral replication. The
 CC present invention relates to synthetic HIV Gag expression cassettes. The
 CC present sequence was cloned and used to generate the expression cassettes
 CC of the present invention. The expression cassettes may be used for the
 CC recombinant expression of HIV Gag-polypeptides which may then be used to
 CC vaccinate against HIV infection and acquired immunodeficiency syndrome
 CC (AIDS).

XX Sequence 2031 BP; 421 A; 707 C; 646 G; 257 T; 0 other;
 SQ

Query Match 1.3%; Score 21; DB 21; Length 2031;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1561 AAGATCTGGCCGAGTACAAG 1581
 |||||
 Db 1318 AAGATCTGGCCGAGTACAAG 1338

RESULT 8
 ACA03520
 ID ACA03520 standard; DNA; 2799 BP.
 AC ACA03520;
 XX
 DT 22-MAY-2003 (first entry)
 DE Synthetic DNA encoding immunogenic HIV peptide #3.
 DE Immunogenic HIV polypeptide; human immunodeficiency virus; HIV;
 KW vaccine; gene therapy; packaging cell line; humoral immune response;
 KW cellular immune response; gene delivery vector; DNA immunisation;
 KW ds.
 XX
 OS Synthetic.
 OS WO2003004657-A1.
 PN
 XX
 XX 16-JAN-2003.
 PD
 XX
 PF 05-JUL-2002; 2002WO-US21421.
 XX
 XX 05-JUL-2001; 2001US-303192P.
 PR 31-AUG-2001; 2001US-316860P.
 PR 16-JAN-2002; 2002US-349728P.
 PR 16-JAN-2002; 2002US-349793P.
 PR 16-JAN-2002; 2002US-349871P.
 XX
 PA (CHIR) CHIRON CORP.
 PI Zur Megede J, Barnett SW, Lian Y;
 XX
 XX WPI; 2003-221602/21.
 DR
 XX
 PT New synthetic polynucleotides encoding antigenic HIV type B and/or type
 PT C polypeptides, useful as immunogenic compositions or vaccines for
 PT generating humoral or cellular immune responses against HIV in a
 PT subject, especially humans -
 XX
 PS Example 1; Fig 8; 262pp; English.
 CC
 CC The invention describes a synthetic polynucleotide encoding 2 or more
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
 CC derived from different HIV subtypes. The polynucleotide is useful for
 CC immunisation, generation of packaging cell lines, or production of HIV
 CC polypeptides. The polynucleotide and its encoded proteins are useful as
 CC immunogenic compositions or vaccines for generating humoral or cellular
 CC immune responses against HIV in a subject, or for inducing neutralising
 CC antibodies against HIV. The gene delivery vector comprising the
 CC polynucleotide is also useful for DNA immunisation of, or for
 CC generating an immune response (e.g. a humoral or cellular immune
 CC response) in, a subject such as a mammal, particularly a human. This
 CC sequence encodes a human immunodeficiency virus immunogenic peptide.
 CC
 XX Sequence 2799 BP; 577 A; 1046 C; 888 G; 288 T; 0 other;

CC antibodies against HIV. The gene delivery vector comprising the
 CC polynucleotide is also useful for DNA immunisation of, or for
 CC generating an immune response (e.g. a humoral or cellular immune
 CC response) in, a subject such as a mammal, particularly a human. This
 CC sequence encodes a human immunodeficiency virus immunogenic peptide.
 CC
 XX Sequence 2799 BP; 577 A; 1046 C; 888 G; 288 T; 0 other;
 SQ

Query Match 1.3%; Score 21; DB 25; Length 2799;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1561 AAGATCTGGCCGAGTACAAG 1581
 |||||
 Db 1318 AAGATCTGGCCGAGTACAAG 1338

RESULT 9
 ACA03583
 ID ACA03583 standard; DNA; 2799 BP.
 AC ACA03583;
 XX
 DT 22-MAY-2003 (first entry)
 DE Synthetic DNA encoding immunogenic HIV peptide #66.
 DE Immunogenic HIV polypeptide; human immunodeficiency virus; HIV;
 KW vaccine; gene therapy; packaging cell line; humoral immune response;
 KW cellular immune response; gene delivery vector; DNA immunisation;
 KW ds.
 XX
 OS Synthetic.
 OS WO2003004657-A1.
 PN
 XX
 XX 16-JAN-2003.
 PD
 XX
 PF 05-JUL-2002; 2002WO-US21421.
 XX
 XX 05-JUL-2001; 2001US-303192P.
 PR 31-AUG-2001; 2001US-316860P.
 PR 16-JAN-2002; 2002US-349728P.
 PR 16-JAN-2002; 2002US-349793P.
 PR 16-JAN-2002; 2002US-349871P.
 XX
 PA (CHIR) CHIRON CORP.
 PI Zur Megede J, Barnett SW, Lian Y;
 XX
 XX WPI; 2003-221602/21.
 DR
 XX
 PT New synthetic polynucleotides encoding antigenic HIV type B and/or type
 PT C polypeptides, useful as immunogenic compositions or vaccines for
 PT generating humoral or cellular immune responses against HIV in a
 PT subject, especially humans -
 XX
 PS Example 1; Fig 71; 262pp; English.
 CC
 CC The invention describes a synthetic polynucleotide encoding 2 or more
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
 CC derived from different HIV subtypes. The polynucleotide is useful for
 CC immunisation, generation of packaging cell lines, or production of HIV
 CC polypeptides. The polynucleotide and its encoded proteins are useful as
 CC immunogenic compositions or vaccines for generating humoral or cellular
 CC immune responses against HIV in a subject, or for inducing neutralising
 CC antibodies against HIV. The gene delivery vector comprising the
 CC polynucleotide is also useful for DNA immunisation of, or for
 CC generating an immune response (e.g. a humoral or cellular immune
 CC response) in, a subject such as a mammal, particularly a human. This
 CC sequence encodes a human immunodeficiency virus immunogenic peptide.
 CC
 XX Sequence 2799 BP; 577 A; 1046 C; 888 G; 288 T; 0 other;

```

Query Match      1.3%; Score 21; DB 25; Length 2799;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCGAGCTACAAG 1581
DB 2602 AAGATCTGGCCGAGCTACAAG 2622

RESULT 10
AAA70472
ID AAA70472 standard; DNA; 4472 BP.
XX
AC AAA70472;
XX
DT 28-NOV-2000 (first entry)
XX
DE HIV bicistronic construct gpl60.modUS4.delV1/V2.Gag.modSF2.
XX
KW HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.
XX
OS Chimeric - Cytomegalovirus.
XX
OS Chimeric - Human immunodeficiency virus type 1.
XX
PN WO200039302-A2.
XX
PD 06-JUL-2000.
XX
PF 30-DEC-1999; 99WO-US31245.
XX
PR 31-DEC-1998; 98US-0114495.
XX
PR 01-DEC-1999; 99US-0168471.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Barnett S, Zur Megede J, Srivastava I, Lian Y, Hartog K, Liu H;
PI Greer C, Selby M, Walker C;
XX
DR WPI; 2000-452400/39.
XX
PT Expression cassettes encoding the human immunodeficiency virus (HIV)
PT Gag-containing polypeptide useful for vaccinating against HIV
PT infections and acquired immunodeficiency syndrome (AIDS) -
XX
PS Claim 21; Fig 63; 391pp; English.
XX
CC The present invention relates to synthetic HIV Gag and Env expression
CC cassettes. The Gag protein of HIV is needed for the assembly of
CC virus-like particles. In addition, the Gag protein is involved in many
CC stages of the HIV life cycle, including assembly, virion maturation after
CC particle release and early post-entry steps in viral replication. The
CC expression cassettes may be used for the recombinant expression of HIV
CC Gag- and Env-polypeptides which may then be used to vaccinate against HIV
CC infection and acquired immunodeficiency syndrome (AIDS). The present
CC sequence is a synthetic construct constructed in the generation of the
CC expression cassettes of the present invention. This construct is
CC bicistronic in that the coding sequences for Env and Gag are under the
CC control of a single CMV promoter and between the two coding sequences an
CC Internal Ribosome Entry Site (IRES) is present.
XX
SQ Sequence 4472 BP; 962 A; 1534 C; 1339 G; 637 T; 0 other;

Query Match      1.3%; Score 21; DB 21; Length 4472;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCGAGCTACAAG 1581
DB 4250 AAGATCTGGCCGAGCTACAAG 4270

RESULT 11
AAA70472
ID AAA70472 standard; DNA; 4472 BP.
XX
AC AAA70472;
XX
DT 28-NOV-2000 (first entry)
XX
DE HIV bicistronic construct gpl60.modUS4.delV1/V2.Gag.modSF2.
XX
KW HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.
XX
OS Chimeric - Cytomegalovirus.
XX
OS Chimeric - Human immunodeficiency virus type 1.
XX
PN WO200039302-A2.
XX
PD 06-JUL-2000.
XX
PF 30-DEC-1999; 99WO-US31245.
XX
PR 31-DEC-1998; 98US-0114495.
XX
PR 01-DEC-1999; 99US-0168471.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Barnett S, Zur Megede J, Srivastava I, Lian Y, Hartog K, Liu H;
PI Greer C, Selby M, Walker C;
XX
DR WPI; 2000-452400/39.
XX
PT Expression cassettes encoding the human immunodeficiency virus (HIV)
PT Gag-containing polypeptide useful for vaccinating against HIV
PT infections and acquired immunodeficiency syndrome (AIDS) -
XX
PS Claim 19; Fig 64; 391pp; English.
XX
CC The present invention relates to synthetic HIV Gag and Env expression
CC cassettes. The Gag protein of HIV is needed for the assembly of
CC virus-like particles. In addition, the Gag protein is involved in many
CC stages of the HIV life cycle, including assembly, virion maturation after
CC particle release and early post-entry steps in viral replication. The
CC expression cassettes may be used for the recombinant expression of HIV
CC Gag- and Env-polypeptides which may then be used to vaccinate against HIV
CC infection and acquired immunodeficiency syndrome (AIDS). The present
CC sequence is a synthetic construct constructed in the generation of the
CC expression cassettes of the present invention. This construct is
CC bicistronic in that the coding sequences for Env and Gag are under the
CC control of a single CMV promoter and between the two coding sequences an
CC Internal Ribosome Entry Site (IRES) is present.
XX
SQ Sequence 4608 BP; 1001 A; 1580 C; 1382 G; 645 T; 0 other;

Query Match      1.3%; Score 21; DB 21; Length 4608;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCGAGCTACAAG 1581
DB 4386 AAGATCTGGCCGAGCTACAAG 4406

RESULT 12
AAA70471
ID AAA70471 standard; DNA; 4689 BP.
XX
AC AAA70471;
XX
DT 28-NOV-2000 (first entry)
XX
DE HIV bicistronic construct gpl60.modUSF162.Gag.modSF2.
XX
KW HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.

```

```
OS Chimeric - Cytomegalovirus.
OS Chimeric - Human immunodeficiency virus type 1.
PN WO200039302-A2.
XX
XX 06-JUL-2000.
XX
XX 30-DEC-1999; 99WO-US31245.
XX
XX 31-DEC-1998; 98US-0114495.
XX
XX 01-DEC-1999; 99US-0168471.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Barnett S., Zur Megede J., Srivastava I., Lian Y., Hartog K., Liu H.;
XX Greer C., Selby M., Walker C.;
XX
XX WPI; 2000-452400/39.
XX
XX Expression cassettes encoding the human immunodeficiency virus (HIV)
XX Gag-containing polypeptide useful for vaccinating against HIV
XX infections and acquired immunodeficiency syndrome (AIDS) -
XX
XX Claim 28; Fig 62; 391pp; English.
XX
XX The present invention relates to synthetic HIV Gag and Env expression
XX cassettes. The Gag protein of HIV is needed for the assembly of
XX virus-like particles. In addition, the Gag protein is involved in many
XX stages of the HIV life cycle, including assembly, virion maturation after
XX particle release and early post-entry steps in viral replication. The
XX expression cassettes may be used for the recombinant expression of HIV
XX Gag- and Env-polypeptides which may then be used to vaccinate against HIV
XX infection and acquired immunodeficiency syndrome (AIDS). The present
XX sequence is a synthetic construct constructed in the generation of the
XX expression cassettes of the present invention. This construct is
XX bicistronic in that the coding sequences for Env and Gag are under the
XX control of a single CMV promoter and between the two coding sequences an
XX Internal Ribosome Entry Site (IRES) is present.
XX
XX Sequence 4689 BP; 1029 A; 1606 C; 1397 G; 657 T; 0 other;
XX
XX Query Match 1.3%; Score 21; DB 21; Length 4689;
XX Best Local Similarity 100.0%; Pred. No. 3.6;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1561 AAGATCTGCCCGAGCTACAAG 1581
XX |||||
XX 4467 AAGATCTGCCCGAGCTACAAG 4487
XX
XX RESULT 13
XX AAA70470
XX ID AAA70470 standard; DNA; 4766 BP.
XX
XX AC AAA70470;
XX
XX XX 28-NOV-2000 (first entry)
XX
XX DE HIV bicistronic construct gp160.modUS4.Gag.modSF2.
XX
XX KW HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.
XX
XX Chimeric - Cytomegalovirus.
XX OS Chimeric - Human immunodeficiency virus type 1.
XX
XX PN WO200039302-A2.
XX
XX 06-JUL-2000.
XX
XX 30-DEC-1999; 99WO-US31245.
XX
XX 31-DEC-1998; 98US-0114495.
XX
XX 01-DEC-1999; 99US-0168471.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Barnett S., Zur Megede J., Srivastava I., Lian Y., Hartog K., Liu H.;
XX Greer C., Selby M., Walker C.;
XX
XX WPI; 2000-452400/39.
XX
XX Expression cassettes encoding the human immunodeficiency virus (HIV)
XX Gag-containing polypeptide useful for vaccinating against HIV
XX infections and acquired immunodeficiency syndrome (AIDS) -
XX
XX Claim 28; Fig 62; 391pp; English.
XX
XX The present invention relates to synthetic HIV Gag and Env expression
XX cassettes. The Gag protein of HIV is needed for the assembly of
XX virus-like particles. In addition, the Gag protein is involved in many
XX stages of the HIV life cycle, including assembly, virion maturation after
XX particle release and early post-entry steps in viral replication. The
XX expression cassettes may be used for the recombinant expression of HIV
XX Gag- and Env-polypeptides which may then be used to vaccinate against HIV
XX infection and acquired immunodeficiency syndrome (AIDS). The present
XX sequence is a synthetic construct constructed in the generation of the
XX expression cassettes of the present invention. This construct is
XX bicistronic in that the coding sequences for Env and Gag are under the
XX control of a single CMV promoter and between the two coding sequences an
XX Internal Ribosome Entry Site (IRES) is present.
XX
XX Sequence 4689 BP; 1029 A; 1606 C; 1397 G; 657 T; 0 other;
XX
XX Query Match 1.3%; Score 21; DB 21; Length 4689;
XX Best Local Similarity 100.0%; Pred. No. 3.6;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1561 AAGATCTGCCCGAGCTACAAG 1581
XX |||||
XX 4467 AAGATCTGCCCGAGCTACAAG 4487
XX
XX RESULT 13
XX AAA70470
XX ID AAA70470 standard; DNA; 4766 BP.
XX
XX AC AAA70470;
XX
XX XX 28-NOV-2000 (first entry)
XX
XX DE HIV bicistronic construct gp160.modUS4.Gag.modSF2.
XX
XX KW HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.
XX
XX Chimeric - Cytomegalovirus.
XX OS Chimeric - Human immunodeficiency virus type 1.
XX
XX PN WO200039302-A2.
XX
XX 06-JUL-2000.
XX
XX 30-DEC-1999; 99WO-US31245.
XX
XX 31-DEC-1998; 98US-0114495.
XX
XX 01-DEC-1999; 99US-0168471.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Barnett S., Zur Megede J., Srivastava I., Lian Y., Hartog K., Liu H.;
XX Greer C., Selby M., Walker C.;
XX
XX WPI; 2000-452400/39.
XX
XX Expression cassettes encoding the human immunodeficiency virus (HIV)
XX Gag-containing polypeptide useful for vaccinating against HIV
XX infections and acquired immunodeficiency syndrome (AIDS) -
XX
XX Claim 27; Fig 61; 391pp; English.
XX
XX The present invention relates to synthetic HIV Gag and Env expression
XX cassettes. The Gag protein of HIV is needed for the assembly of
XX virus-like particles. In addition, the Gag protein is involved in many
XX stages of the HIV life cycle, including assembly, virion maturation after
XX particle release and early post-entry steps in viral replication. The
XX expression cassettes may be used for the recombinant expression of HIV
XX Gag- and Env-polypeptides which may then be used to vaccinate against HIV
XX infection and acquired immunodeficiency syndrome (AIDS). The present
XX sequence is a synthetic construct constructed in the generation of the
XX expression cassettes of the present invention. This construct is
XX bicistronic in that the coding sequences for Env and Gag are under the
XX control of a single CMV promoter and between the two coding sequences an
XX Internal Ribosome Entry Site (IRES) is present.
XX
XX Sequence 4766 BP; 1051 A; 1638 C; 1409 G; 668 T; 0 other;
XX
XX Query Match 1.3%; Score 21; DB 21; Length 4766;
XX Best Local Similarity 100.0%; Pred. No. 3.6;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1561 AAGATCTGCCCGAGCTACAAG 1581
XX |||||
XX 4544 AAGATCTGCCCGAGCTACAAG 4564
XX
XX RESULT 14
XX ACA03519
XX ID ACA03519 standard; DNA; 4773 BP.
XX
XX AC ACA03519;
XX
XX XX 22-MAY-2003 (first entry)
XX
XX DE Synthetic DNA encoding immunogenic HIV peptide #2.
XX
XX KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV;
XX vaccine; gene therapy; packaging cell line; humoral immune response;
XX cellular immune response; gene delivery vector; DNA immunisation;
XX ds.
XX
XX OS Synthetic.
XX
XX PN WO2003004657-A1.
XX
XX XX 16-JAN-2003.
XX
XX PF 05-JUL-2002; 2002WO-US21421.
XX
XX PR 05-JUL-2001; 2001US-303192P.
XX
XX PR 31-AUG-2001; 2001US-316860P.
XX
XX PR 16-JAN-2002; 2002US-349728P.
XX
XX PR 16-JAN-2002; 2002US-349793P.
XX
XX PR 16-JAN-2002; 2002US-349871P.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Zur Megede J., Barnett SW., Lian Y.;
XX WPI; 2003-221602/21.
XX
XX
```


XX New synthetic polynucleotides encoding antigenic HIV type B and/or type
 PT C polypeptides, useful as immunogenic compositions or vaccines for
 PT generating humoral or cellular immune responses against HIV in a
 PT subject, especially humans -
 XX
 PS Example 1; Fig 7; 262pp; English.
 XX
 CC The invention describes a synthetic polynucleotide encoding 2 or more
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
 CC derived from different HIV subtypes. The polynucleotide is useful for
 CC immunisation, generation of packaging cell lines, or production of HIV
 CC polypeptides. The polynucleotide and its encoded proteins are useful as
 CC immunogenic compositions or vaccines for generating humoral or cellular
 CC immune responses against HIV in a subject, or for inducing neutralising
 CC antibodies against HIV. The gene delivery vector comprising the
 CC polynucleotide is also useful for DNA immunisation of, or for
 CC generating an immune response (e.g. a humoral or cellular immune
 CC response) in, a subject such as a mammal, particularly a human. This
 CC sequence encodes a human immunodeficiency virus immunogenic peptide.
 XX
 SQ Sequence 4773 BP; 1035 A; 1698 C; 1506 G; 534 T; 0 other;

Query Match 1.3%; Score 21; DB 25; Length 4773;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
 |||||
 Db 1318 AAGATCTGGCCCGAGCTACAAG 1338

RESULT 15
 ACA03585
 ID ACA03585 standard; DNA; 4773 BP.
 XX
 AC ACA03585;
 XX
 DT 22-MAY-2003 (first entry)
 XX
 DE Synthetic DNA encoding immunogenic HIV peptide #68.
 XX
 KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV;
 KW vaccine; gene therapy; packaging cell line; humoral immune response;
 KW cellular immune response; gene delivery vector; DNA immunisation;
 KW ds.
 XX
 OS Synthetic.
 OS
 XX
 PN WO2003004657-A1.
 XX
 PD 16-JAN-2003.
 XX
 PF 05-JUL-2002; 2002WO-US21421.
 XX
 PR 05-JUL-2001; 2001US-303192P.
 PR 31-AUG-2001; 2001US-316860P.
 PR 16-JAN-2002; 2002US-349728P.
 PR 16-JAN-2002; 2002US-349733P.
 PR 16-JAN-2002; 2002US-349871P.
 XX

FA (CHIR) CHIRON CORP.
 XX
 XX Zur Megede J, Barnett SW, Lian Y;
 PI
 XX
 DR WPI; 2003-221602/21.
 XX

XX New synthetic polynucleotides encoding antigenic HIV type B and/or type
 PT C polypeptides, useful as immunogenic compositions or vaccines for
 PT generating humoral or cellular immune responses against HIV in a
 PT subject, especially humans -
 XX

PS Example 1; Fig 73; 262pp; English.

XX The invention describes a synthetic polynucleotide encoding 2 or more
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
 CC derived from different HIV subtypes. The polynucleotide is useful for
 CC immunisation, generation of packaging cell lines, or production of HIV
 CC polypeptides. The polynucleotide and its encoded proteins are useful as
 CC immunogenic compositions or vaccines for generating humoral or cellular
 CC immune responses against HIV in a subject, or for inducing neutralising
 CC antibodies against HIV. The gene delivery vector comprising the
 CC polynucleotide is also useful for DNA immunisation of, or for
 CC generating an immune response (e.g. a humoral or cellular immune
 CC response) in, a subject such as a mammal, particularly a human. This
 CC sequence encodes a human immunodeficiency virus immunogenic peptide.
 XX

SQ Sequence 4773 BP; 1035 A; 1698 C; 1506 G; 534 T; 0 other;

Query Match 1.3%; Score 21; DB 25; Length 4773;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
 |||||
 Db 2602 AAGATCTGGCCCGAGCTACAAG 2622

Search completed: November 21, 2003, 14:49:12
 Job time : 401 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 18:05:53 ; Search time 2610 Seconds
(without alignments)
5168.193 Million cell updates/sec

Title: US-09-712-338-2
Perfect score: 2979
Sequence: 1 MRGYEFLSVLPLVAASWALP.....HTQSSVPLPTATSMSSVGMA 555

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-O=/cgn2_1/USPTO_spool_p/US09712338/runat_17112003_170522_11545/app_query.fasta_1.711
-DB=EST_QPMP=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -FAST=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CNG 1.1 3549 @runat_17112003_170522_11545 -NCPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_nam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rtd.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	512	17.2	588	9	AW672518	LGI_360_F
2	409	13.7	1049	29	CNS06XNN	AL420285 T3 end of
3	405.5	13.6	648	10	BE777028	BE777028 MY-23-G-0
4	405.5	13.6	653	14	CD258740	CD258740 pIMY023ag
5	403.5	13.5	907	14	CD456151	CD456151 Fq03_10g0
6	385.5	12.9	1002	29	CNS062OK	AL422586 T3 end of
7	384	12.9	1054	29	CNS06NN7	AL406985 T7 end of
8	375.5	12.6	886	29	CNS06G80	AL410350 T7 end of
9	354	11.9	1005	29	CNS06EAK	AL394866 T7 end of
10	353	11.8	3268	11	AK032197	AK032197 Mus muscu
11	352.5	11.8	963	29	CNS0625E	AL421896 T3 end of
12	349.5	11.7	2493	11	AK019861	AK019861 Mus muscu
13	345	11.6	721	14	CA747625	CA747625 GAP13D07
14	333	11.2	1581	11	AF217508	AF217508 Homo sapi
15	332.5	11.2	712	14	CD045123	CD045123 PsH017XN
16	327.5	11.2	729	13	BQ110016	BQ110016 VD0102C06
17	327.5	11.0	685	13	BW251033	BW251033 BW251033
18	327.5	11.0	700	13	BW246209	BW246209 BW246209
19	326	10.9	1492	11	AK017087	AK017087 Mus muscu
20	324.5	10.9	876	14	CD384438	CD384438 FTMW09611
21	322	10.8	652	9	AV991679	AV991679 AV991679
22	321	10.8	879	29	CNS0750D	AL429491 clone BAO
23	320.5	10.8	675	13	BW274166	BW274166 BW274166
24	320.5	10.8	675	13	BW302983	BW302983 BW302983
25	320.5	10.8	680	13	BW208298	BW208298 BW208298
26	320.5	10.8	684	13	BW248944	BW248944 BW248944
27	320.5	10.8	685	13	BW194502	BW194502 BW194502
28	320.5	10.8	686	13	BW260398	BW260398 BW260398
29	320.5	10.8	689	13	BW046109	BW046109 BW046109
30	320.5	10.8	692	13	BW194721	BW194721 BW194721
31	320.5	10.8	699	13	BW035379	BW035379 BW035379
32	319.5	10.7	566	13	BW276795	BW276795 BW276795
33	319.5	10.7	656	13	BW276796	BW276796 BW276796
34	319.5	10.7	702	13	BW212011	BW212011 BW212011
35	319.5	10.7	762	13	BQ165650	BQ165650 EST611519
36	319	10.7	485	12	B1200653	B1200653 O1a06fs.1
37	318	10.7	676	14	CD044351	CD044351 PsH014XA
38	318	10.7	700	12	B1749802	B1749802 F902_05d0
39	316.5	10.6	682	13	BW255378	BW255378 BW255378
40	312	10.5	708	14	CD213567	CD213567 HSI_41_F1
41	311.5	10.5	638	13	BW282449	BW282449 BW282449
42	309.5	10.4	656	13	BW292110	BW292110 BW292110
43	308	10.3	804	14	CB644508	CB644508 OSNEB05M
44	307.5	10.3	580	13	BW197413	BW197413 BW197413
45	307	10.3	637	14	CA352124	CA352124 623325 NC

ALIGNMENTS

RESULT 1
AW672518
LOCUS
DEFINITION
LGI_360_F03.b1_A002 Light Grown 1 (LGI) Sorghum bicolor cDNA, mRNA
588 bp mRNA linear EST 19-JUL-2000
ACCESSION
AW672518
VERSION
AW672518.1
GI:7536439
KEYWORDS
EST.
SOURCE
Sorghum bicolor (sorghum)
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Sorghum.

```

REFERENCE
AUTHORS      1 (bases 1 to 588)
TITLE        Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
LOCUS        An EST database from Sorghum: light-grown seedlings
JOURNAL      Unpublished
COMMENT      Contact: Cordonnier-Pratt MM
              Laboratory for Genomics and Bioinformatics
              The University of Georgia, Department of Plant Biology
              Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
              Tel: 706 542 1860
              Fax: 706 583 0210
              Email: mmpratt@uga.edu
              Sequences have been trimmed to exclude PolyA, vector and regions
              below Phred quality 16. The threshold for highest quality sequence
              is 20.
              Seq primer: JEN REV
              POLYA-No.
              Location/Qualifiers
                1..588
                /organism="Sorghum bicolor"
                /mol_type="mRNA"
                /db_xref="taxon:4558"
                /clone_lib="Light Grown_1 (LG1)"
                /note="Organ: 10- to 14-day-old light-grown (greenhouse)
                seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI
                ; The library was made from poly-A RNA in the cloning
                vector lambda ZAP II. Clones to be sequenced were
                prepared by mass excision."
BASE COUNT   145 a 151 c 152 g 140 t
ORIGIN
Alignment Scores:
Pred. No.:    2,29e-45      Length:    588
Score:         512.00      Matches:    96
Percent Similarity: 71.15%      Conservative: 15
Best Local Similarity: 61.54%      Mismatches: 41
Query Match:    17.19%      Indels:    4
DB:             9          Gaps:      2

US-09-712-338-2 (1-555) x AW672518 (1-588)
QY 32 LeuProLysAsnProThrGlyValLysThrLeuThrThrAlaAsnValThrIleArg 51
DB 128 CTGCCAAGGAAGTACATGACATACAGACGGCTGACGCCAACCAATCTACTATCGA 187
QY 52 TyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSer 71
DB 188 TACAGAACCTT-----GGTATCTGTGAGACGACGCGCTGGTGTACAGACTACTCA 238
QY 72 GlyTyrValAspThrSerProGluSerHisThrPhePheTrpPheGluAlaArgHis 91
DB 239 GGATACGTGGATCTCGACACCAATGCCACGTCTTCTTCTGGTTCTTCAGTCTAGGAAC 298
QY 92 AsnProGluThrAlaProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeu 111
DB 299 AATCCCGCATCTGATCCGTCTTGTGGTGAATGGCGGTCCCGATCAGATTCGCTG 358
QY 112 IleGlyLeuPheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIle 131
DB 359 AITGGGCTCTTGAAGAAACGGTCCATCATGATGATAGTATATCTCAGCGGTGAGTAC 418
QY 132 AsnProHisSerTrpAsnGluValSerAsnLeuPheLeuSerGlnProLeuGlyVal 151
DB 419 AACCCCTCAGCTGGAAACATGCTCTCAACATCTATATATTTCCACGGCGGTGCGCACG 478
QY 152 GlyPheSerTyrSerThrValAspGlySerIleAsnProValThrGlyValValGlu 171
DB 479 GGATTCAGTACACGAGCAAGCGCTCGCGGATTTGATTCACATCAGGAGAAATCCGT 538
QY 172 ---AsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAsp 186
DB 539 TACAACCTCGAGTGAATACCGGCTGATGGCGGTACCGGCTCTTGGAT 586

```

```

RESULT 2
CNS06XWN      1049 bp      DNA      linear      GSS 06-JUL-2001
LOCUS         T3 end of clone AY0AA004D06 of library AY0AA from strain CBS 6340
DEFINITION    of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION     AL420285
VERSION       AL420285.1 GI:12203469
KEYWORDS      GSS.
SOURCE        Kluyveromyces thermotolerans
ORGANISM      Kluyveromyces thermotolerans
REFERENCE     1 (bases 1 to 1049)
AUTHORS       Souciet,J.L., Aglie,M., Artiguenave,F., Blandin,G.,
              Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
              de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorente,B.,
              Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
              Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
              Wincker,P. and Weissenbach,J.
              Genomic exploration of the hemiascomycetous yeasts: 1. A set of
              yeast species for molecular evolution studies
              FEMS Lett. 487 (1), 3-12 (2000)
TITLE         Kluyveromyces thermotolerans
JOURNAL       FEMS Lett. 487 (1), 61-65 (2000)
MEDLINE       20584720
PUBMED        11152885
REFERENCE     2 (bases 1 to 1049)
AUTHORS       Malpertuy,A., Liorente,B., Blandin,G., Artiguenave,F., Wincker,P.
              and Dujon,B.
              Genomic exploration of the hemiascomycetous yeasts: 10.
              Kluyveromyces thermotolerans
              FEMS Lett. 487 (1), 61-65 (2000)
TITLE         Kluyveromyces thermotolerans
JOURNAL       FEMS Lett. 487 (1), 61-65 (2000)
MEDLINE       20584720
PUBMED        11152885
REFERENCE     3 (bases 1 to 1049)
AUTHORS       Direct Submission
TITLE         Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
              2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail:
              segref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
              This GSS is part of a random genomic sequencing program of thirteen
              yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
              exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
              Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Pichia
              lactis var. lactis, Kluyveromyces marxianus var. hansenii, Pichia
              angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
              Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
              5 kb were prepared and both extremities were sequenced. See
              keywords for description of this sequence and for the sequence of
              the other extremity of this insert.
COMMENT
FEATURES
source
  1..1049
  /organism="Kluyveromyces thermotolerans"
  /mol_type="genomic DNA"
  /strain="CBS 6340"
  /db_xref="taxon:4916"
  /clone="AY0AA004D06"
  /clone_lib="AY0AA"
  /note="end: T3"
misc_feature
  <82..>1020
  /note="similar to Saccharomyces cerevisiae ORF YBR139w [
  strong similarity to carboxypeptidase Y]"
  /evidence=not_experimental
misc_feature
  <172..>1017
  /note="similar to Saccharomyces cerevisiae ORF YMR297w [
  PC1; carboxypeptidase Y, serine-type protease]"
  /evidence=not_experimental
BASE COUNT   262 a 253 c 252 g 280 t
ORIGIN
Alignment Scores:
Pred. No.:    9.24e-34      Length:    1049
Score:         409.00      Matches:    114
Percent Similarity: 43.90%      Conservative: 37
Best Local Similarity: 33.14%      Mismatches: 137

```

Query Match: 13.73% Indels: 56
DB: 29 Gaps: 10

US-09-712-338-2 (1-555) x CNS06XWN (1-1049)

QY	19	LeuProGlySerThrProAlaSerValGly-----ArgArgGlnLeuProLysAsnPro	36
Db	109	GTCCCAAAATACCGTCAGCTCATCAAAAATGTCGGTGAAGTCTGCAGCAGCTGCTC	168
QY	37	ThrGlyValLysThrLeuThrThrAlaAsnAsnValThrIleArgTy-LysGluProgly	56
Db	169	ACGATAAAGTCTGAGCTTGATGACACCTACTCGATGAGGTAAAGAGAGTGGACCCITCG	228
QY	57	AlaGluGlyValCysGluThrThrProGlyValLysSerTy-SerGlyTyValAspThr	76
Db	229	AAGCTGGGTGTC-----GATTCTGTCAAAGCAGTACTCGGGGTATCTGGACTAC	276
QY	77	SerProGluSerHisThrPheThrPhePheGluAlaArgHisAsnProGluThrAla	96
Db	277	GAGGATCCAACACCTCTTTTACTGGCATTTGAATCCAGAAGACCCTCAATGAC	336
QY	97	ProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu	116
Db	337	CCGTTATTCTATGGCTGAACGGTGTGCCAGGTCTCTCTTTCACGGGTATTITTC	396
QY	117	GluLeuGlyProCysHisValAsnSerThrPheAspAspTy-TyrIleAsnProHisSerTrp	136
Db	397	GAACITGGCCCCCTCTTCGTCGGCCCTTGAGCTAAAGCCCCGTTCGCAATCCGTA	456
QY	137	AsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTy-Ser	156
Db	457	ATAACAANTGCTACTGTATATCTTGAGCAGCCCTTAGCGGTGGTTTTCTCTATGGT	516
QY	157	AspThrValAspGlySerIleAsnProValThrGlyValGluAsnSerSer-PheAla	176
Db	517	GAC-----	519
QY	177	GlyValGlnGlyArgTy-ProThrIleAspAlaThrLeuIleAspThrAsnLeuAla	196
Db	520	-----GAGAGAGTTGGCTCCATCAACGGCAGCA	546
QY	197	AlaGluAlaAAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSer	216
Db	547	GGAAAGAGTGTCTCATATCTTGAGCTGTTTTTCCAGAGATCCCA-----	594
QY	217	ArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyGlyHisTySerGlyPro	236
Db	595	CAGTTTAGATCTAACGACTTCCATTCGAGGTGAATCATACGCTGGCCATTACATCCA	654
QY	237	AlaPhePheAsnHisPheTy-GluGlnAsnGluArgIleAlaAsnGlySerValAsnGly	256
Db	655	-----GAGATTGCCCATCAAAATCGCTGCTTCATGAATCTGAC	693
QY	257	ValGlnLeuAsnPheAsnSerLeudylelleAsnGlyllelleAspGluAlalleGln	276
Db	694	AAACCTTCAATCTCACCTCCATCATGATTGCCACCGGGATACAGATTCTTAGTCCAG	753
QY	277	AlaProTyTyProGluPheAlaValAsnAsnThrTyTyGlylleLysAlaValAsnGlu	296
Db	754	TATGACTACTATGAGCCCATGCGGTGGCAGGGGT---GGTTACAAAGCCGTG---ATA	807
QY	297	ThrValTyAsnTyMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeulle	316
Db	808	ACTGAGGAAGATGGCCCAAAATGAGAACCAAAATGGCTCGC---TCGACAGCTTTGAAC	864
QY	317	SerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyTyAlaLeuCysAlaGluAla	336
Db	865	AATGCATGCTACTCTCTAGTTCACG-----TTCGGTTGATTGGCAGCT	909
QY	337	ThrAsnMetCysArgAspAsnValGluGlyProTyTyTyAlaPheAlaGlyArgGlyVal	356
Db	910	GGCGCATACTGT---GAGAAATATGCCCATGAGTGTCTTACACTAAACAGGCCCTGAACGT	966

357 TyrAspIleArg 360
|||||
967 TATGACATACGC 978

RESULT 3
BE777028
LOCUS BE777028
DEFINITION MY-23-G-01 PinfestansMY Phytophthora infestans cDNA, mRNA sequence.
ACCESSION BE777028
VERSION BE777028.1 GI:10230683
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.

REFERENCE
1 (bases 1 to 648)
Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.
Initial assessment of gene diversity for the oomycete pathogen
Phytophthora infestans based on expressed sequences
Fungal Genet. Biol. 28 (2), 94-106 (1999)
20056376
PUBMED 10587472
CONTACT: Govers F
Laboratory of Phytopathology
Wageningen University
Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands
Tel: 31 317 483 138
Fax: 31 317 483 412
Email: Francine.Govers@medew.fyto.wau.nl.

FEATURES
Location/Qualifiers
1..648
/organism="Phytophthora infestans"
/mol_type="mRNA"
/strain="DDR7602, A1 mating type"
/db_xref="taxon:4787"
dev_stage="4-week old vegetative, non-sporulating mycelium in synthetic medium"
lab_host="E. coli, strain DH5-alpha"
clone_lib="PinfestansMY"
note=Vector: pSPOR1; Site_1: SalI; Site_2: NotI; Total RNA was isolated from mycelium of P. infestans DDR7602 cultured for 4 weeks in synthetic medium. EST clones were named by their position in the microtiter plate, preceded by the prefix MY (for mycelial) and the successive number of the microtiter plate (e.g. MY-06-A-04)."

BASE COUNT 161 a 174 c 168 g 145 t
ORIGIN

Alignment Scores:
Pred. No.: 1,15e-33 Length: 648
Score: 405.50 Matches: 86
Percent Similarity: 50.63 Conservative: 35
Best Local Similarity: 35.98 Mismatches: 77
Query Match: 13.61 Indels: 41
DB: 10 Gaps: 6

US-09-712-338-2 (1-555) x BE777028 (1-648)

QY	58	GluGlyValCysGluThrThrProGlyValLysSerTy-SerGlyTy-	ValAsp 75
Db	48	GAGGATTTTGTGATCGACC-----AAGCAGTTGAGTCGTCTTCAAGATAACA	98
QY	76	ThrSerProGluSerHisThrPhePhePhePheGluAlaArgHisAsnProGluThr	95
Db	99	GGCTCCAAGTCGAGAACTACTTTTACTGTTCTTTCGATACGCGGAGTCCCTCGACC	158
QY	96	AlaProLthrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPhe	115
Db	159	GATCGCTGATCATCTGGCTTACTGGCGTCTCTGGATCGAGCTCCATCTTGGCTTTACTG	218
QY	116	GluGluLeuGlyProCysHisValAsnSerThrPheAspAspTy-TyrIleAsnProHisSer	135
Db	219	CAGGAAAAAGCCCGTCTCGGTGAACGACGACTTCTCACTCAAGAAAGAAATCCGTATTCG	278

```

QY 136 TrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyr 155
Db 279 TGGACGAGCGTGGCCAAATGTCATGTGGATCGACCGCGTGGTGGCTTACGTAC 338
QY 156 SerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPhe 175
Db 339 GGAGAC-----ATGTTCCACCTTCCTGCAAGAAATCTTCAAGGCCTTGCCA----- 344
QY 176 AlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeu 195
Db 345 -----AGCGTGAATACACACGTCAGACAGAGAGGTGGGCGACAC----- 386
QY 196 AlaAlaGluAlaAlaTrpGluLeuLeuGlnGlyPheLeuSerGlyLeuProSerLeuAsp 215
Db 387 -----ATGTTCCACCTTCCTGCAAGAAATCTTCAAGGCCTTGCCA----- 425
QY 216 SerArgValGlnSerIleAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGly 235
Db 426 ---GAGTATCAAAAACCTCCCTTCTACGCTTCGGAGAGAGTTACGCGCGCCACCTACGTTG 482
QY 236 ProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsn 255
Db 483 CCGCGGATCGCTCAGACAACTTTACTGCAACACAGCAA-----AAGGAGGA 530
QY 256 GlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIle 275
Db 531 CCTGTGGAAATTAATTTGAAGGCTTTGGAATTGGGACGGCTCACCCGACCTGAAAGTG 590
QY 276 GlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleValAlaVal 294
Db 591 CAGTCAAGTACTACCCGACATGGCGCTACAAACAATACATACGCGCTCAAGGCTGTG 647

```

```

RESULT 4
CD258740 653 bp mRNA linear EST 23-MAY-2003
LOCUS piMY023aG01r_213792 Wageningen/U.Maryland Phytophthora infestans
DEFINITION EST Project Phytophthora infestans cDNA clone IMY023G01 5, mRNA
sequence.
ACCESSION CD258740.1 GI:31046567
VERSION CD258740
KEYWORDS Phytophthora infestans (potato late blight agent)
SOURCE Phytophthora infestans
ORGANISM Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.

```

```

REFERENCE 1 (bases 1 to 653)
AUTHORS Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.
TITLE Initial assessment of gene diversity for the oomycete pathogen
Phytophthora infestans based on expressed sequences
JOURNAL Fungal Genet. Biol. 28 (2), 94-106 (1999)
MEDLINE 20056376
PUBMED 10587472
COMMENT Contact: Tyler B
Tyler lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmtyle@vt.edu
PCR primers
FORWARD: M13 reverse 17mer at 5' end
BACKWARD: M13 reverse 17mer at 5' end
Plate: 023 row: G column: 01
Seq primer: M13 reverse 17mer at 5' end
High quality sequence stop: 653.
Location/Qualifiers
1. 653
/organism="Phytophthora infestans"
/mol_type="mRNA"
/db_xref="taxon:4787"
/clone="IMY023G01"
/tissue_type="mycelium"
/cell_line="DDR7602"

```

```

FEATURES
source
1. 653
/organism="Phytophthora infestans"
/mol_type="mRNA"
/db_xref="taxon:4787"
/clone="IMY023G01"
/tissue_type="mycelium"
/cell_line="DDR7602"

```

```

/dev_stage="mycelium"
/clone_lib="Wageningen/U.Maryland Phytophthora infestans
EST Project"
/note="Vector: pCMV-SPORT6.1; Site_1: SalI; Site_2: NotI"
BASE COUNT 161 a 176 c 170 g 146 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1.16e-33 Length: 653
Score: 405.50 Matches: 86
Percent Similarity: 50.63% Conservative: 35
Best Local Similarity: 35.98% Mismatches: 77
Query Match: 13.61% Indels: 41
DB: 14 Gaps: 6

```

```

US-09-712-338-2 (1-555) x CD258740 (1-653)
QY 58 GluGlyValCysGluThrThrProGlyValIleValSerTyrSerGlyTyr-----ValAsp 75
Db 53 GAGGATTTTGTGATCTCGACC-----AAGCAGTTGAGTGGTCTTCAAGATAACA 103
QY 76 ThrSerProGluSerHisThrPheThrPhePheGluAlaArgHisAsnProGluThr 95
Db 104 GGCCTCAAGTCGAGAACTACTTTTACTGGTTCTTTCGAATCACCGCGGAGTCCCTCGACC 163
QY 96 AlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPhe 115
Db 164 GATCGCTGATCATCTGGCTTACTGCGGTCTCTGATGACAGTCTCATCTTGTCTTACGTG 223
QY 116 GluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSer 135
Db 224 CAGGAAACCGGCCCGTCTCCGTGAAACGACGACTTGTCTCACTCAAGAGAAATCCGTATTCTG 283
QY 136 TrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyr 155
Db 284 TGGAAACGAGCGTGCCTAATGTCATGTGATCGACACCGCGTGGTGGCTTCAAGTAC 343
QY 156 SerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPhe 175
Db 344 GGAGAC-----ATGTTCCACCTTCCTGCAAGAAATCTTCAAGGCCTTGCCA----- 349
QY 176 AlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeu 195
Db 350 -----AGCGTGAATACACACGTCAGACAGAGAGGTGGGCGACAC----- 391
QY 196 AlaAlaGluAlaAlaTrpGluLeuLeuGlnGlyPheLeuSerGlyLeuProSerLeuAsp 215
Db 392 -----ATGTTCCACCTTCCTGCAAGAAATCTTCAAGGCCTTGCCA----- 430
QY 216 SerArgValGlnSerIleAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGly 235
Db 431 ---GAGTATCAAAAACCTCCCTTCTACGCTTCGGAGAGAGTTACCGCGGACCTACGTG 487
QY 236 ProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsn 255
Db 488 CCGCGGATCGCTCAGACAACTTTACTGCAACACCAACCA-----AAGGAGGA 535
QY 256 GlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIle 275
Db 536 CCGTGTGAAATTAATTTGAAGGCTTTGGAATTGGGACGGCTCACCCGACCTTGAAGTG 595
QY 276 GlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleValAlaVal 294
Db 596 CAGTCAAGTACTACCCGACATGGCGCTACAAACAATACATACGCGCTCAAGGCTGTG 652

```

```

RESULT 5
CD456151/c 907 bp mRNA linear EST 03-JUN-2003
LOCUS Fg03_10907_R
DEFINITION Fg03_AAFc_ECORC_Fusarium graminearum mycelium trichothecene_product
ion Gibberella zeae cDNA clone Fg03_10907, mRNA sequence.
ACCESSION CD456151
VERSION CD456151.1 GI:31370891

```

242	PheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn	Phe	261
906	TTTGAACAGCAGAAATGCTGCTATGATCTGTGTCATCAAGGGCAGAA	GATCAACSTC	847
262	AsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyr	TyrPro	281
846	GTAGTCTTGGCATCAACAATGGATGGATTGATCTGGCCAACCATG	CAGAGATTACATC	787
282	GluPheAlaValAsnAsnThrTyrGlyIleIleValAlaAsnGluThrValTyr	AsnTyr	301
786	GAATACGCAGCAATAACACCTTACAAA--AAGYTGATCAGCTCAACT	-----CAGTAC	736
302	MethLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr	CysLysGln	321
735	AGCAAGTACCTAAACACCTACAATCAAAAGGTGTGCTGTTCGCSAAG	TGCCAGCCAGGCC	676
322	ThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn	MetCysArg	341
675	TCACAGCAACGAGCT-GCT-----TGTGGCAATGGCAGATGTCGAGT	CTTGCACT	629

MEDLINE
 PUBMED
 20584722
 11152887
 REFERENCE
 3
 (bases 1 to 1002)
 AUTHORS
 Genoscope.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
1..1002
/organism="Kluyveromyces marxianus"
/mol_type="genomic DNA"
/strain="CBS 712"
/variety="marxianus"
/db_xref="taxon:4911"
/clone="AZ0AA003B09"
/clone_lib="AZ0AA"
/note="end : T3"
complement(<8..>994)
/note="similar to Saccharomyces cerevisiae ORF YMR297w [
PRC1 : carboxypeptidase Y, serine-type protease]"
evidence=not experimental
complement(<11..>778)
/note="similar to Saccharomyces cerevisiae ORF YBR139w [
strong similarity to carboxypeptidase]"
evidence=not experimental
BASE COUNT 267 a 259 c 240 g 235 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 3-21e-31 Length: 1002
Score: 385.50 Matches: 102
Percent Similarity: 42.59% Conservative: 36
Best Local Similarity: 31.48% Mismatches: 123
Query Match: 12.94% Indels: 63
DB: 29 Gaps: 11

US-09-712-338-2 (1-555) x CNS06ZOK (1-1002)
QY 36 ProThrGlyValLysThrLeuThrThrAlaAsnValThrIle----- 50
DB 826 CCAGAGAGCTTTGTCTACGGCTAGCGTCAACAGCAGCAAGTCCGGAACCTACCGAGTTG 767
QY 51 -----ArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyVal 67
DB 766 CGTGTGAAGAGATCAAGGACCTTAAGTCTTGGGCTG-----GACCCCGAGCTG 716
QY 68 LysSerTyrSerGlyTyrValAspThrSerProGlu---SerHisThrPhePheTrpPhe 86
DB 715 AAGCAATACCTCGGTTACTTGGAGCTTGAAGACGAGCAGCAGCACTTCTTACTGTTC 656
QY 87 PheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyPro 106
DB 655 TTTGAGTCCAGAAACACCCCAAGACACCTCTATCTTGTGGTGAACGGTGGGCCA 596
QY 107 GlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsnSerThr 126
DB 595 GGGTGTTCCTTTTGACTGTTTGTCTTCAATTGGTCCATCTCCATGGCGAGAG 536
QY 127 PheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSer 146
DB 535 GTGAAGCAATTTACACCCACACTCTTGGAAACAGCAACGCTCCGTTATCTTCTTGAC 476
QY 147 GlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProVal 166
DB 475 CAGCCAGTCAACGTTGGTTACTCTTACTCTTCACTCA---GAAGGTGTCTCAAACTGTT 419
QY 167 ThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAsp 186
DB 418 GCT----- 416

QY 187 AlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGluIleuGlnGly 206
DB 415 -----CGAGTAAGGACGCTGATCGCTTCTTGCATTTG 383
QY 207 PheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeuTrp 226
DB 382 TTTCTCCAGCAATTTCCCGAGTACGCCTCT-----GGCCAAGATTTCCACATTGCT 332
QY 227 ThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPhe-----TyrGlu 244
DB 331 GGTGAATCTTATGCGGGCCACTACATCCCTGTGTTGCTACAGAGATTTTGTGCGACCCA 272
QY 245 GlnAsnGluAsnGlyAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeu 264
DB 271 CAAGAGAAACGT-----TCTTCAACTGACCTCTGTT 239
QY 265 GlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAla 284
DB 238 TTGATCGGTACGGTTTGACCGACCCCTTGACCCAGTACCCTGAGTACGAGCCTATGGCT 179
QY 285 ValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPhe 304
DB 178 TGTGTCTCAAGTGTGTGAGCCATCTGTGTTGGACGAGGAGCAATGTGAAGCATG----- 125
QY 305 AlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArg 324
DB 124 TTGGAAGCTTACCCAGAG---TGCTAAACTTGATTCATCTTGCTACGAA----- 77
QY 325 ThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnVal 344
DB 76 -----TAGCAATCTGTTGCTCACTGTGCGCAGCTCCATCTACTGTAAACAACGCTCAA 23
QY 345 GluGlyProTyr 348
DB 22 ATGGGTCCATAC 11
RESULT 7
CNS06NN7
LOCUS
DEFINITION
T7 end of clone AU0AA015E08 of library AU0AA from strain CBS 3082
of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION AL406985 1054 bp DNA linear GSS 04-JUL-2001
VERSION AL406985.1 GI:12171743
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri
ORGANISM Saccharomyces kluyveri
REFERENCE 1 (bases 1 to 1054)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekla,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 1054)
AUTHORS Neuveglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
Gallardin,C. and Casaregola,S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 9.
Saccharomyces kluyveri
JOURNAL FEBS Lett. 487 (1), 56-60 (2000)
MEDLINE 20584719
PUBMED 11152884
REFERENCE 3 (bases 1 to 1054)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,


```

PUBMED 11152892
REFERENCE 3 (bases 1 to 986)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    Location/Qualifiers
        1..986
            /organism="Yarrowia lipolytica"
            /mol_type="genomic DNA"
            /strain="CLIB 89"
            /db_xref="taxon:4952"
            /clone="AW0AA03H06"
            /clone_lib="AW0AA"
            /note="end : T7"
            <54..>938
            /note="similar to Saccharomyces cerevisiae ORF YMR297w [
            PRCl : carboxypeptidase Y, serine-type protease ]"
            /evidence=not_experimental
            <54..>920
            /note="similar to Saccharomyces cerevisiae ORF YBR139w [
            strong similarity to carboxypeptidase ]"
            /evidence=not_experimental
            257 a 251 c 249 g 227 t 2 others
            misc_feature
            misc_feature
BASE COUNT 257 a 251 c 249 g 227 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 3..89e-30 Length: 986
Score: 375.50 Matches: 102
Percent Similarity: 40.44% Conservative: 44
Best Local Similarity: 28.25% Mismatches: 148
Query Match: 12.60% Indels: 67
DB: 29 Gaps: 11
US-09-712-338-2 (1-555) x CNS06Q80 (1-986)
QY 38 GlyValThrLeuThr-----ThrAlaAsnValThrIleArgTyrLysGluPro 55
Db 9 GGAATCAAGATGGCTGCTGACGAGTTTCAGATGGCACATTCGGGTCACTGATCCC 68
QY 56 GlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAsp 75
Db 69 GCAGACTAGGAGTG-----GATTCGTCAGCAACATTCAGGCTACTTTGAT 116
QY 76 ThrSerProGluSer---HisThrPhePheTyrPheGluAlaArgHisAsnProGlu 94
Db 117 TTCGAACAGAAATACAAACACCTGTTTCTTGGTTCTTCTGAGTCGGGAAACGACCCCAA 176
QY 95 ThrAlaProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 114
Db 177 ACGAGACCTGTAGTCCTGTGGATCAGCGAGGACCCGGTGTCTTCCATAAAGGGGATG 236
QY 115 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis 134
Db 237 TTCTTTGAGATGGCTCAGCAAGGTGCAACCCGAGCTCAACTTGTGGACAATCCCTAC 296
QY 135 SerTyrAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer 154
Db 297 GCATGGAATTAACAACGCTCCGTCATCTACTCGATCAGCGCGGTGATCTGGCTACTCA 356
QY 155 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer 174

```

```

Db 357 TACTCTCTCA----- 365
QY 175 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsn 194
Db 366 -----GATGAACACAGAGTCAACTCCACGAGA 392
QY 195 LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu 214
Db 393 CAAGCAGCCACAGGACGTCATCGGTTTCTTAAACAAGTTCTTTGAAGTTTACCCC----- 446
QY 215 AspSerArgValGlnSerLysAspPheSerLeuTyrThrGluSerTyrGlyGlyHisTyr 234
Db 447 -----GAATACGCCGAGCTGGATTTCACGTGGCTGGAGAGTCGTATGCTGGCCATTAC 500
QY 235 GlyProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerVal 254
Db 501 ATTCTGCTATCGCTACCGAGATCCAAAGTCCAAAGGAGAG------ 542
QY 255 AsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIle-----AsnGlyIleIleAsp 272
Db 543 -----AACTACGAGTTGGCTCGGTGCTGATTGGAAACGGAGTCCACAGAC 587
QY 273 GluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLys 292
Db 588 ACAAAAGACCCAGTGCCTTACTTCCAGGCCATGGCTTGTGGAGAGGAGATACCCCTTCT 647
QY 293 AlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 312
Db 648 GTTCTGTCGGAAGACACCTGCCAGCAGATGAAGGATGCC-----GTTGGAGACTGC 698
QY 313 GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu 332
Db 699 CAGAAGCTCATTTGAACACTGCTGGGAAACCCCAACCAAGCTCAGTGTACGCC--- 755
QY 333 CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAla 352
Db 756 -----ACATGACTACTGCTCTCTCTGTCGAGCCCTACATCGGTGAAGGA 806
QY 353 GlyArgGlyValTyrAspIleArgHisPro-----TyrAspAspPro 366
Db 807 GGACGAAATGCTTACGATATCCGAGACGCTGTGCTCCAGACTGGCTGTTCACGAGCAG 866
QY 367 ThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGly 386
Db 867 AAC-----ATTGCCATTGAGTCTTACCTCAAGATGAACATGTCAGACTGTTCTTGGT 920
QY 387 Val 387
Db 921 GTC 923
RESULT 9
CNS06EAK 1005 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AR0AA020A09 of library AR0AA from strain CBS 732 of
DEFINITION Zygosaccharomyces rouxii, genomic survey sequence.
ACCESSION AL394866
VERSION AL394866.1 GI:12146085
SOURCE GSS.
ORGANISM Zygosaccharomyces rouxii
Zygosaccharomyces rouxii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.
REFERENCE 1 (bases 1 to 1005)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekai,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
GENOMIC exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711

```

11152876
2 (bases 1 to 1005)
de Montigny, J., Straub, M., Potier, S., Tekalia, F., Dujon, B.,
Wincker, P., Artiguenave, F. and Souciet, J.
Genomic exploration of the hemiascomycetous yeasts: 8.
Zygosaccharomyces rouxii
FEMS Lett. 487 (1), 52-55 (2000)
20584718
11152883
3 (bases 1 to 1005)
Genoscope.
Direct Submission
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
Location/Qualifiers
1..1005
/organism="Zygosaccharomyces rouxii"
/mol_type="genomic DNA"
/strain="CBS 732"
/db_xref="taxon:4956"
/clone="AR0AA020A09"
/clone_lib="AR0AA"
/note="end : T7"
misc_feature
<3..>985
/note="similar to Saccharomyces cerevisiae ORF YBR139w [
strong similarity to carboxypeptidase]
1 putative frameshift(s)"
/evidence=not_experimental
misc_feature
<3..>982
/note="similar to Saccharomyces cerevisiae ORF YMR297w [
PRL1 ; carboxypeptidase Y, serine-type protease]
1 putative frameshift(s)"
/evidence=not_experimental
BASE COUNT 284 a 187 c 231 g 302 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 8.95e-28 Length: 1005
Score: 354.00 Matches: 117
Percent Similarity: 41.62% Conservative: 42
Best Local Similarity: 30.63% Mismatches: 155
Query Match: 11.88% Indels: 70
DB: 29 Gaps: 11
US-09-712-338-2 (1-555) x CNS06EAK (1-1005)
QY 144 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 163
Db 1 TTCTTGGACACACCGTAGT-GTGGTTTTTCATATGGTGAT----- 41
QY 164 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 183
Db 41 ----- 41
QY 184 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIle 203
Db -----GAAAAGGTTTACCAATACCAGAGCTGCTGGTAGGACGCTACACATCTC 89
QY 204 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerIlysAspPhe 223
Db ----- 90 TTGGAATTATCTTCGAAAAGTTCCACATTTG-----AGATCGCATGACTTTC 137

QY 224 SerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyr 243
Db 138 CACATTGCTGGTGAATCTCTATGCTGGACATTACATTCTCAGATTGCCCATGAATCGTT 197
QY 244 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 263
Db 198 ATCGAATCTCT-----GGAAGAACTTTTGACCTTAACCTCT 233
QY 264 LeuGlyIleIleAsnGlyIleAspGluAlaIleGlnAlaProTyrTyrProGluPhe 283
Db 234 GTTCTTATCGGCAATGGTATTACCGATTCCCTGATTCAAATGATTACTATCAGCCAATG 293
QY 284 AlaVal---AsnAsnThrTyrGlyIleIleValAsnGluThrValTyrAsnTyrMet 302
Db 294 GCTTGGCGGTGAAGTGGATATCTCACTTTAACTATGATGAAGATTGTTACGATGGAG 353
QY 303 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysGlyGlnThr 322
Db 354 AAAAATACAAATCGT-----TGCCGTACGTTGAACCGTATCTGTTACGGTACA 401
QY 323 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp 342
Db 402 AAATCTAAC-----ATTGCATGTGTAGTCTGCTACTGCTATCTGTGAATCT 446
QY 343 AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro 362
Db 447 GTTACTGTGGGCACATTCCAAGAAAAAACTGGTGAATGTCTATGATATCAGGGTCTCT 506
QY 363 TyrAspAspProThrPro-----ProSerTyrTyrAsnLysPheLeu 376
Db 507 TGTGAGGATAATGATAGTGCACCTGCTACTTTGGTATGAATATGTTGATGATCATG 566
QY 377 AlaLysAspSerValMetAspAlaIleGlyValAsnIle---AsnTyrThrGlnSerAsn 395
Db 567 AACCAAGATACGTACAGGACGCGTTGGGGTCAGATGTACACAATATACCGTTGTAAC 626
QY 396 AsnAspValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGlu 415
Db 627 GATCAAGTATCTTTGGGATTGGCCCTAACTGGGGATGGTGCCAAAGCCT---TTCCAACA 683
QY 416 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAsp 435
Db 684 TATGTACAGAACTGGTGGATTAAACATTCACATCTTGTCTACCGTGTGTATAAGAT 743
QY 436 TyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAla 455
Db 744 TACATTTGTAACCTGGTTGGGTAAAGCTTGGTCTGATAAATAGATTGGAGATACCGT 803
QY 456 AlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----AsnGly 470
Db 804 GAAAAGTACGAATCT-----CTACCTCTAAAGCGCTGGAAAAAGTCAAAGTACTGTT 854
QY 471 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
Db 855 GAAAATTTGGTGAAGTTAAAC-TACGGCGCTTTGACCTTCTTAAGGATTTACGATGCT 913
QY 491 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 510
Db 914 GGACACATGTTCTTATGACCAGCCTGAAGCGCTCTTGAATGTTGAAGTGTGAACGACTGATT 973
QY 511 PheGly 512
Db 974 ACAGGG 979
RESULT 10
AK032197
LOCUS
DEFINITION
Mus musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430411I23 product:protective protein for
beta-galactosidase, full insert sequence.
AK032197
ACCESSION
VERSION
AK032197.1 GI:26328022
KEYWORDS
HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374

3 Ito, T., Itoh, M., Aizawa, K., Kagitani, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system—384-format
sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., De Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Marzella, J., Mombarts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, T.,
Wysshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 695-690 (2001)
21085660

5 The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
11217851

6 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Takagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Toyama, A., Toya, T., Yasunishi, A.,
Yasuda, T.

```

QY 67 ValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheTrpPhe 86
Db 169 TTCGGCAATACCTCGGCTACCTCAGAGATCGAGTCCAGACATCCACTACTGTTT 228
QY 87 PheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyPro 106
Db 229 GTGGAGTCGAGAACAGACAGCCAAAGAACAGCCCGCTGGTCTTTGGCTTAACGGGGTCCC 288
QY 107 GlySerAspSerLeuLeuGlyLeuPheGluGluLeuGlyProCysHisValAsnSer--- 125
Db 289 GGTGTCAGCTCGCTCGATGGCTGCTTACAGACAGCGCCCTTCTGATCCAGCCAGAT 348
QY 126 -----ThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeu 143
Db 349 GGTGTCCACCTGGAG-----TACACCCCTATGCTTGGACCTGATGCCAAGCGCTG 402
QY 144 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 163
Db 403 TATATCGAGTCCCGCAGCTGGGTGGGCTTCTCTACTCGGAT----- 444
QY 164 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 193
Db 444 ----- 444
QY 184 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaIleTrpGluIle 203
Db 445 -----GACAAGATGTACGTGACCAATGACACAGAGTGGCGAGAACAAATTAAGACC 498
QY 204 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 223
Db 499 CTTAAAGACTCTTCGGCCCTCTTCCG-----GAATACAGGACAAC 540
QY 224 SerLeuTrpThr-----GluSerTyrGlyHisTyrGlyProAlaPheAsnHis 241
Db 541 AAACITTTCTGACAGAGAGACTATGTGGCATCTACATCCCCACC----- 588
QY 242 PheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPhe 261
Db 589 -----TTGGCTGTACTGGTCATGCAGATCTAGCATGAACTT 627
QY 262 AsnSerLeuGlyIleLeuAsnGlyIle-----AspGluAlaIle 275
Db 628 CAGGGCTGGCTGTGGCAATGACTTGCCTCTATGAGCAGAACGACAACTCCCTGGTC 687
QY 276 GlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsn 295
Db 688 TACTTTGGCTTACTACCATGGCTTCTGGGAACAGACTTTGGACTTCACTGCAGACCCAC 747
QY 296 GluThrValTyrAsnTyrMetLysPhe---AlaAsnGlnMetProAsnGlyCysGlnAsp 314
Db 748 TGTGCTGGCTCAGAACAGTGAATCTTCTATGACAAACAGACCCAGAGTGTGAACAAT 807
QY 315 LeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAla 334
Db 808 CTCCTGGAAGTGTCTCGAATTTGGCAATCTGGCTCAACATCTACAATCTCTATGCT 867
QY 335 GluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArg 354
Db 868 CCG-----TGTGCTGGTGGGTGCCGCG----- 891
QY 355 GlyValTyrAspIleArgHisProTyrAspAsp----- 365
Db 892 -----AGACATACATATGAGACACACTTGTAGTCCAGATTTTGGCAAC 936
QY 365 ----- 365
QY 937 ATCTTCACTCGCTGCCACTTAAGCGGAGATTTCTTGAGSCACTGATCGTTCTGGGAC 996
QY 366 -----ProThrProProSerTyrTyr-----Asn 373
Db 997 AAGGTAGCTTGGATCCTCCCTGCACCAACACACAGCCCTTCCAACTACTCTCAACAAC 1056

```

```

QY 374 LysPheLeuAlaLys-----AspSerVal-----MetAspAla 384
Db 1057 CCTATGTTTCGGAGAGCTTCCACATCCCGAGTCGTCGCCCGCTGGGACATGTGCAAC 1116
QY 385 IleGlyValAsnIleAsnTyrThr-----GlnSerAsnAsnAspValTyrTyrAla 401
Db 1117 TTCTTGTGAATTTACAGTACCGCCCTCTACCAAGCATGACATCCCACTAC----- 1170
QY 402 PheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIleLeu 421
Db 1171 -----CTGAAGCTGCTCAGT 1185
QY 422 AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 441
Db 1186 TCACAGAAATACCATCTGCTCTACACGGAGATGTGGACATGCGCTGCAACTTCATG 1245
QY 442 GlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 461
Db 1246 GCGCATGAGTGGTGTGGATTGCTCAACAGAGATGGAGGTGCAGCGCCCG----- 1299
QY 462 GlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGlu-----Thr 476
Db 1300 -----CCCTGGCTAGTGGACTACGGGAGCGGAGAACAGGTAGCTGTTTCGTG 1350
QY 477 ArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyr 496
Db 1351 AAGAGTGTTCACACATCCTTCCTCACCATCAAGGTGCGCGACACATGGTCCCAAC 1410
QY 497 TyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 508
Db 1411 GACAAGCTCGAGCTGCTTTTACCATGTTCTCGAGG 1446

```

RESULT 11

```

CNS0625E          963 bp      DNA      linear      GSS 07-JUL-2001
LOCUS             T3 end of clone AY0AA015G04 of library AYCAA from strain CBS 6340
DEFINITION        of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION         AL421896
VERSION           AL421896.1 GI:12205091
KEYWORDS          GSS.
SOURCE            Kluyveromyces thermotolerans
ORGANISM          Kluyveromyces thermotolerans
REFERENCE         Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS           Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
                  1 (bases 1 to 963)
                  Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
                  Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
                  de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
                  Malbertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
                  Saurin,W., Tekaiia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
                  Wincker,P. and Weissenbach,J.
                  Genomic exploration of the hemiascomycetous yeasts: 1. A set of
                  yeast species for molecular evolution studies
                  FEBS Lett. 487 (1), 3-12 (2000)
                  20584711
                  PUBLISHED
                  11152876
REFERENCE         2 (bases 1 to 963)
AUTHORS           Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.
                  and Dujon,B.
                  Genomic exploration of the hemiascomycetous yeasts: 10.
                  Kluyveromyces thermotolerans
                  FEBS Lett. 487 (1), 61-65 (2000)
                  20584720
                  PUBLISHED
                  11152885
REFERENCE         3 (bases 1 to 963)
AUTHORS           Genoscope.
TITLE             Direct Submission
JOURNAL           Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
MEDLINE           2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
PUBLISHED         seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
AUTHORS           This GSS is part of a random genomic sequencing program of thirteen
TITLE             yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
JOURNAL           exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
COMMENT

```

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source
1..963
Location/Qualifiers
/organism="Kluyveromyces thermotolerans"
/mol_type="genomic DNA"
/strain="CBS 6340"
/db_xref="taxon:4916"
/clone="AY0AA015G04"
/clone_lib="AY0AA"
/note="end: T3"
misc_feature
270..>962
/note="similar to Saccharomyces cerevisiae ORF YBR139W [strong similarity to carboxypeptidase]"
evidence=not experimental
BASE COUNT 242 a 231 c 217 g 272 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 1.23e-27 Length: 963
Score: 352.50 Matches: 88
Percent Similarity: 45.08% Conservative: 31
Best Local Similarity: 33.33% Mismatches: 98
Query Match: 11.83% Indels: 47
DB: 29 Gaps: 5

US-09-712-338-2 (1-555) x CNS06Z5E (1-963)

QY	19	LeuProGlySerHisThrProAlaSerValGly-----ArgArgGlnLeuProHisAsnPro	36
Db	297	GNCCCAAAATACCGTGCAGTCTATCAAAATGCGTGAAGTCTGCAGGATCGTC	356
QY	37	ThrGlyValLysThrLeuThrThrAlaAsnValThrIleArgTyrLysGluProGly	56
Db	357	ACGATAAAGCTCGAGCTTGATGACACTACTCGATGAGGTAAAGAGATGACCCCTCG	416
QY	57	AlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThr	76
Db	417	AGCTGGGTGTC-----GATTCTGCAAGCAGTACTCGGGGTATCTGGACTAC	464
QY	77	SerProGluSerHisThrPheThrPhePheGluAlaArgHisAsnProGluThrAla	96
Db	465	GAGGATTCACCAACACTCTTTTACTGGCATTTGATCCAGAAACGACCCACTCATGAC	524
QY	97	ProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu	116
Db	525	CCCGTTATTCTATGGCTGAACGGTGTCTCTCTTTCACGGGTATTATTTTC	584
QY	117	GluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrp	136
Db	585	GAACCTGGCCCTCTTCGGTCGCCCTGAGCTAAACGCCGCTTCGCAATCGTACTCTGG	644
QY	137	AsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTrpSer	156
Db	645	AATAACAATGCTACTGCTATATTCTTGGAGCAGCCCTTAGCGGTTCCTATGTTTCT	704
QY	157	AspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAla	176
Db	705	GAC-----	707
QY	177	GlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAla	196
Db	708	-----GAGAGATTCCCTCCACTACGACGCA	734
QY	197	AlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSer	216
Db	735	GGAAGGAGTGTTCTATATTCTTGGAGTCTGTTTTCACAGAGTCCCA	782

QY	217	ArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyPro	236
Db	783	CAGTTTAGATCTACAGACTTCATATTCAGGTGAATCATACGCTGGCATTTACATCCCA	842
QY	237	AlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGly	256
Db	843	-----GAGTTGCCCATCAATCGCTGCTTCATCAATCTGAC	881
QY	257	ValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGln	276
Db	882	AAACCTTCATCTCACCTCCATCATGATTGCAACGGGATAACAGATTCTTTAGTCCAG	941
QY	277	AlaProTyrTyr	280
Db	942	TATGAMTACTAT	953
RESULT	12		
LOCUS	AK019861	2493 bp	mRNA linear HTC 05-DEC-2002
DEFINITION	Mus musculus 11 days pregnant adult female ovary and uterus cDNA, RIKEN full-length enriched library, clone:503140F07		
	product:protective protein for beta-galactosidase, full insert sequence.		
ACCESSION	AK019861		
VERSION	AK019861.1	GI:12860233	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349636		
AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, H., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159		
TITLE	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861		
JOURNAL	4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakata, I., Saio, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, J., Fleischmann, W., Gaasterland, T., Gissi, C., King, P., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staib, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,		


```

Db      821 CCCACC-----TTGGCTGTACTGTCATGCAG 847
QY      256 GlyValGlnLeuAsnPheAsnSerLeuGlylleleAsnGlyllele----- 271
Db      848 GATCCTAGCATGAATCTTCAGGGGCTGGTGGGCAATAGGACTTGCCTCTATGAGCAG 907
QY      272 -----AspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyr 289
Db      908 AACGCAACTCCTCGTCTACTTTCCTACTACCATGCGCTTCTCGGGAACAGACTTTGG 967
QY      290 GlylleLysAlaValAsnGluThrValTyrAsnTyrMetLysPhe---AlaAsnGlnMet 308
Db      968 ACTTCAGTCGACACCCTGCTGGCTCAGACAGTGAATCTTATGACACAAAGAC 1027
QY      309 ProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAla 328
Db      1028 CCAGAGTGTGTAACAATCTCTCGGAAGTGTCTCGAATTGTGGCAAAATCTCGCCTCAAC 1087
QY      329 AspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyr 348
Db      1088 ATCTCAATCTCTATGCTCCG-----TGTGCTGGTGGGTGCCCGC----- 1129
QY      349 TyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAsp----- 365
Db      1130 -----AGACATAGATATGAGCACACACTGTGA 1156
QY      365 ----- 365
Db      1157 GTCCAGGATTTTGGCAACATCTTCACTCGCTGCCACCTTAAGCGGAGATTTCTCGAGGCA 1216
QY      366 -----ProThrProPro 369
Db      1217 CTGATCGGTTCGGGACAAAGTACGCTTGGATTCCTTCCTGCAACCAACACCGCCCT 1276
QY      370 SerTyrTyr-----AsnLysPheLeuAlaLys-----AspSerVal--- 381
Db      1277 TCAACTACTCAACAAACCCCTATGTTCGGAAGGCTCTCCACATCCCCAGTGTGCTGCC 1336
QY      382 -----MetAspAlaIleGlyValAsnIleAsnTyrThr-----GlnSerAsn 395
Db      1337 CGCTGGGACATGTGCAACTTCTTGTGTAATTTACAGTACCGCGCTCTTACCAAGCATG 1396
QY      396 AsnAspValTyrTyrAlaPheGlnThrGlyAspPheValTyrProAsnPheIleGlu 415
Db      1397 AACTCCAGTAC----- 1408
QY      416 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAsp 435
Db      1409 ---CTGAAGCTGCTCAGTTACAGAAATACCAGATCCTCTCTACAAACGAGATGGAC 1465
QY      436 TyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAla 455
Db      1466 ATGGCGTCAACTTCATGGCGGATGATGGTGTGTGGATTTCGCTCAACAGAGATGGAG 1525
QY      456 AlaGlnPheArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGlu 475
Db      1526 GTGACGCGCGG-----CCCTGGCTAGTGGACTACCGGAGAGCGGAGAA 1570
QY      476 -----ThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
Db      1571 CAGGTAGCTGGTTTCTGTGAAGGAGTGTTCACACATCACCTTCTCTCACCACCAAGGTGCC 1630
QY      491 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 508
Db      1631 GGACACATGTCCTCCCAACGACACAGCCTCGAGCTGCTTTTACCATGTCTCGAGG 1684

RESULT 13
CA747625 721 bp mRNA linear EST 14-JAN-2003
LOCUS GAP13D07 Aspergillus flavus Lambda Zap cDNA Library Aspergillus
DEFINITION flavus cDNA, mRNA sequence.
ACCESSION CA747625
VERSION CA747625.1 GI:27735664

```

```

KEYWORDS EST.
SOURCE Aspergillus flavus
ORGANISM Aspergillus flavus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 721)
AUTHORS Obrian,G.R., Fakhoury,A.M. and Payne,G.A.
TITLE Identification of Genes Differentially Expressed During Aflatoxin
Biosynthesis in Aspergillus flavus and Aspergillus parasiticus
JOURNAL Unpublished (2003)
COMMENT Contact: Payne, GA
Department of Plant Pathology
North Carolina State University
Box 7616, Gardner Hall, Raleigh, NC 27695, USA
Tel: (919) 515-6994
Fax: (919) 515-7716
Email: Gary_Payne@NCSTU.
FEATURES
source
1..721
/organism="Aspergillus flavus"
/mol_type="mRNA"
/strain="3357"
/db_xref="taxon:5059"
/clone_lib="Aspergillus flavus Lambda Zap cDNA Library"
/note="Vector: pBluescript; Log phase of aflatoxin
production"
BASE COUNT 161 a 159 c 224 g 176 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 5 45e-27 Length: 721
Score: 345.00 Matches: 66
Percent Similarity: 61.54% Conservative: 30
Best Local Similarity: 42.31% Mismatches: 60
Query Match: 11.58% Indels: 0
DB: 14 Gaps: 0
US-09-712-338-2 (1-555) x CA747625 (1-721)
QY 387 ValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyrAlaPheGlnGlnThrGly 406
Db 8 GTCCCTGTCACTTCCACGAGTCCGCCAGTGGTCCACAAACAGTTTGACAAAGACCTAC 67
QY 407 AspPheValTyrProAsnPheIleGluAspLeuGluIleLeuAlaLeuProValArg 426
Db 68 GACATCACCCGCGCGGATTCCTCGACTCGATAGCCCTCTCGACAGCGCGTCAAG 127
QY 427 ValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaVal 446
Db 128 GTACATGATGATGACGCGACCGGACTACGCTGTACTGGATCGCGCGGAGAAAGCC 187
QY 447 SerLeuAlaAlaAsnTyrSerGlnAlaGlnPheArgSerAlaGlyTyrThrProLeu 466
Db 188 AGCCTCGAGTACCGTACTCGCGGCGCGAGAGTTCGCAACACCCGGGTATATCCATT 247
QY 467 LysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArg 486
Db 249 GTGACGTCGCAAGGATCAAGGGTATGACACGCTCAACTAGGTAAATATAGCTTCACGCG 307
QY 487 ValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPhe 506
Db 308 GTCTACCGCTGGTTCATGAGTGGCGGCTTATCAGCGGTTTCGGCGGTATGAGATCTTC 367
QY 507 AsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLysIleTrpProSerTyr 526
Db 368 ATGCGTGTCTACGCTGGATCGGATATCTCTACTGGAGAGATAGTATTACGGGGAGTTT 427
QY 527 LysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValPro 542
Db 428 AAGACGAGTGGACCGGAGGATATCTTGGCATATTAGAAATGTGCTCCA 475

RESULT 14
AF217508

```


QY 470 lyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluA 490
 Db 1301 -----ATCCGGCAAGTGGGTGACTTCCATCAGGTAATATTTCGAGGTG 1343
 QY 490 leGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrI 510
 Db 1344 GAGGACATATTATACCTATGACCGCTCTGAGAGCTTTGACATGATTAATCGATTCA 1403
 QY 510 lePhe-----GlyTTPASP 514
 Db 1404 TTTATGAAAAGATGGAT 1423

RESULT 15
 LOCUS CD045123
 DEFINITION 712 bp mRNA linear EST 09-MAY-2003
 psHB017N05f 183846 psHB: Infected hypocotyl soybean host. 48 hrs
 post infection Phytophthora sojae cDNA clone SHB017N05 5, mRNA
 sequence.

ACCESSION CD045123
 VERSION CD045123.1 GI:30498716
 KEYWORDS EST.
 SOURCE Phytophthora sojae
 ORGANISM Phytophthora sojae
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 Phytophthora.

REFERENCE 1 (bases 1 to 712)
 AUTHORS Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
 TITLE USDA-IPAFS: Expression of Phytophthora sojae genes during infection
 and propagation

JOURNAL Unpublished
 COMMENT Contact: Tyler B
 Tyler lab
 VBI

1880 Pratt Dr., Blacksburg, VA 24061, USA
 Tel: 540-231-7319
 Email: bmtylex@vt.edu
 PCR Primers
 FORWARD: BK reverse
 Plate: 017 row: N column: 05
 Seg primer: BK reverse
 High quality sequence stop: 712.
 Location/Qualifiers

FEATURES
 source
 1..712
 /organism="Phytophthora sojae"
 /mol_type="mRNA"
 /db_xref="taxon:67593"
 /clone="SHB017N05"
 /tissue_type="infected host tissue"
 /cell_line="F6497"
 /dev_stage="48 hour post infection"
 /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs
 post infection"
 /note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI;
 USDA-IPAFS: Expression of Phytophthora sojae genes during
 infection and propagation."
 BASE COUNT 154 a 224 c 218 g 116 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,25e-25 Length: 712
 Score: 332.50 Matches: 82
 Percent Similarity: 44.65% Conservative: 39
 Best Local Similarity: 30.26% Mismatches: 97
 Query Match: 11.16% Indels: 53
 DB: 14 Gaps: 8

US-09-712-338-2 (1-555) x CD045123 (1-712)

QY 28 GlyArgGlnLeuProLysAsnProThrGlyValLysThrLeuThrAlaAsnAsn 47
 Db 42 GGCTGGCGGCCCATCGAGAACAGTGGCGGCTCGCTACCTGCTACTGTCGGGCGAC 101

QY 48 ValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyVal 67
 Db 102 ACAGTC-----ATGTGGCACC-----GCA 122
 QY 68 LysSerTyrSerGlyTyrValAspThr-----SerProGluSerHisThrPhePheTyr 85
 Db 123 ACCAAGAGCGGGCTACGTCAAGTCGCGCACAGCAGACCATCTTCTTACTGG 182
 QY 86 PhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTyrLeuAsnGlyGly 105
 Db 183 TTCTTCGAGTCCCGCAGAGCTCCTGCTACCGACCCCTTGGTCTCTGTTGAGCGTGGC 242
 QY 106 ProGlySerAspSerLeuIleGlyLeuGluLeuGlyProCysHisValAsnSer 125
 Db 243 CCGCGAGTTTCAGGCTCATGACGCTCTTGACCGAGAACGCCCGCTTCGTTAAGGAG 302
 QY 126 ThrPheAspAspTyrIleAsnProHisSerTyrAsnGluValSerAsnLeuLeuPheLeu 145
 Db 303 GATATGTCGAGCGGCGCAACCCCACTCGTGGACTCGGAGCCCAACGTCTCTGCTG 362
 QY 146 SerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnPro 165
 Db 363 GACCAAGCCACGAAAGTGGGTACTCGTACGCGAGCCCGCAGATCCGAC-----CACGAC 419
 QY 166 ValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIle 185
 Db 420 GAGAAGGAGCTCCAGGAGAAC-----GTCTACGGGTCTCTACAG 458
 QY 186 AspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaThrGluIleLeuGln 205
 Db 441 -----GTCTACGGGTCTCTACAG 458
 QY 206 GlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeu 225
 Db 459 GGGTTTCTGGACAAACACCCCGGAGCTT-----CAAGACGAGCGGCTGTTCCTC 506
 QY 226 TrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGln 245
 Db 507 GCAGGCGAGAGCTACGCGAGCCACATCCCGAGCGCGCGCACAGATCCACCGCGAG 566
 QY 246 AsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGly 265
 Db 567 AACAAA-----GGCGGAAGTCAAGCGCTAAACCTCCAGGSCATCGCC 611
 QY 266 IleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaVal 285
 Db 612 ATCGGGAACGGCTTGACCAACACCGCTCGTGCAGAGCGGACACGGGCTGGACATG---GTC 668
 QY 286 AsnAsnThrTyrGlyIleLysAlaValAsnGlu 296
 Db 669 AACAACTCGTACGGCGTCAAGTGTATGGACGAC 701

Search completed: November 21, 2003, 20:19:18
 Job time : 2626 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus p2n model

Run on: November 21, 2003, 18:07:18 ; Search time 474 Seconds
(without alignments)
3826.695 Million cell updates/sec

Title: US-09-712-338-2
Perfect score: 2979
Sequence: 1 MRGVEFLVSLPLVAASWALP.....HTQSSVPLPLTATSSVGVMA 555

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Xgapop 10.0, Xgapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09712338/runat_17112003_170523_11592/app.query.fasta_1.711
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnnp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09712338 @CGN 1 1 107 arunat 17112003 170523_11592
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCGAPOP=6 -FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB ID	Description
1	588.5	19.8	2218	14	US-10-206-619-1	Sequence 1, Appli

RESULT 1

US-10-206-619-1
; Sequence 1, Application US/10206619
; Publication No. US20030108908A1
; GENERAL INFORMATION:
; APPLICANT: Korea Research Institute of Bioscience and Biotechnology
; APPLICANT: Dong Kook Pharmaceutical Co.
; APPLICANT: Rhee, Sangki
; APPLICANT: Choi, Euisung
; APPLICANT: Kang, Hyunah
; APPLICANT: Sohn, Junghoon
; APPLICANT: Bae, Junghoon
; APPLICANT: Kim, Moowong
; APPLICANT: Agaphonov, Michasel
; TITLE OF INVENTION: Hansenula polymorpha mutants and process for the preparation of
; FILE REFERENCE: 4220-116 US
; CURRENT APPLICATION NUMBER: US/10/206,619
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US/09/674,617
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: PCT/KR00/00173
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0

ALIGNMENTS

2	475	15.9	1551	10	US-09-938-842A-683	Sequence 683, App
3	366.5	12.3	1949	12	US-09-971-429B-33	Sequence 33, Appli
4	361	12.1	1551	13	US-10-084-018-4	Sequence 4, Appli
5	359	12.1	1697	11	US-09-796-753-39	Sequence 39, Appli
6	359	12.1	2076	10	US-09-909-320-163	Sequence 163, App
7	359	12.1	2076	10	US-09-909-088B-163	Sequence 163, App
8	359	12.1	2076	10	US-09-905-291A-163	Sequence 163, App
9	359	12.1	2076	10	US-09-902-853-163	Sequence 163, App
10	359	12.1	2076	10	US-09-907-824-163	Sequence 163, App
11	359	12.1	2076	10	US-09-907-841-163	Sequence 163, App
12	359	12.1	2076	11	US-09-904-011-163	Sequence 163, App
13	359	12.1	2076	11	US-09-906-742-163	Sequence 163, App
14	359	12.1	2076	11	US-09-906-838-163	Sequence 163, App
15	359	12.1	2076	11	US-09-907-613-163	Sequence 163, App
16	359	12.1	2076	11	US-09-907-942-163	Sequence 163, App
17	359	12.1	2076	11	US-09-904-859-163	Sequence 163, App
18	359	12.1	2076	11	US-09-909-204-163	Sequence 163, App
19	359	12.1	2076	11	US-09-904-820-163	Sequence 163, App
20	359	12.1	2076	11	US-09-904-786-163	Sequence 163, App
21	359	12.1	2076	11	US-09-906-646-163	Sequence 163, App
22	359	12.1	2076	11	US-09-906-700-163	Sequence 163, App
23	359	12.1	2076	11	US-09-903-786-163	Sequence 163, App
24	359	12.1	2076	11	US-09-902-903-163	Sequence 163, App
25	359	12.1	2076	11	US-09-903-749A-163	Sequence 163, App
26	359	12.1	2076	11	US-09-904-119-163	Sequence 163, App
27	359	12.1	2076	11	US-09-904-956-163	Sequence 163, App
28	359	12.1	2076	11	US-09-902-736-163	Sequence 163, App
29	359	12.1	2076	11	US-09-907-794-163	Sequence 163, App
30	359	12.1	2076	11	US-09-903-943-163	Sequence 163, App
31	359	12.1	2076	11	US-09-904-462-163	Sequence 163, App
32	359	12.1	2076	11	US-09-907-925-163	Sequence 163, App
33	359	12.1	2076	11	US-09-902-692-163	Sequence 163, App
34	359	12.1	2076	11	US-09-903-520-163	Sequence 163, App
35	359	12.1	2076	11	US-09-905-056-163	Sequence 163, App
36	359	12.1	2076	11	US-09-909-064-163	Sequence 163, App
37	359	12.1	2076	11	US-09-904-553-163	Sequence 163, App
38	359	12.1	2076	11	US-09-905-381-163	Sequence 163, App
39	359	12.1	2076	11	US-09-905-088-163	Sequence 163, App
40	359	12.1	2076	11	US-09-907-575-163	Sequence 163, App
41	359	12.1	2076	11	US-09-905-075-163	Sequence 163, App
42	359	12.1	2076	11	US-09-902-759-163	Sequence 163, App
43	359	12.1	2076	11	US-09-902-634-163	Sequence 163, App
44	359	12.1	2076	11	US-09-907-979-163	Sequence 163, App
45	359	12.1	2076	11	US-09-907-979-163	Sequence 163, App

```

; SEQ ID NO 1
; LENGTH: 2218
; TYPE: DNA
; ORGANISM: Hansenula polymorpha DL1
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: U67174
; DATABASE ENTRY DATE: 1996-08-17
; RELEVANT RESIDUES: (1)..(2218)
US-10-206-619-1

Alignment Scores:
Pred. No.: 6,74e-59 Length: 2218
Score: 588.50 Matches: 169
Percent Similarity: 45.61% Conservative: 70
Best Local Similarity: 32.25% Mismatches: 200
Query Match: 19.75% Indels: 86
DB: 14 Gaps: 20

US-09-712-338-2 (1-555) x US-10-206-619-1 (1-2218)
QY 19 LeuProGlySerThrProAlaSerValGlyArgArgGln-----Leu 32
DB 640 ATACCTCAGAAACATGAGCTCACTAGTCAGCCCGGCGAGTCAAATTACAAAGAACCTG 699
QY 33 ProLysAsnProThrGlyValLysThrLeuThrAlaAsnValThrIleArgTyr 52
DB 700 CCGATTTTGGGACITCAATGTCTCAATGAGAAGTTC-TCCAACTTACAAGCTGAGGGTT 758
QY 53 Lys-----GluProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyr 70
DB 759 AAGAAACCGACCGCCGGGACATGGGACTGGACACACA-----AGACAGTAC 806
QY 71 SerGlyTyrValAspThrSerProGlu-----SerHisThrPhePhePheGluAla 89
DB 807 TCTGATACCTGGATGTGAGACGACGACGACGACGACGACGACGACGACGACGACGAC 866
QY 90 ArgHisAsnProGluThrAlaProLethrLeuThrLeuAsnGlyGlyProGlySerAsp 109
DB 867 AGAATGACCCGGTCAACGACCTGTGATCTGTGCTCAACGGTGTGCTGACGAGTCTCT 926
QY 110 SerLeuIleGlyLeuPheGluGluGluGluGluGluGluGluGluGluGluGluGlu 129
DB 927 TCTTTGACGGAATGCTTTTGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 986
QY 130 TyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeu 149
DB 987 ATCAACACCCATATTCGTGGATTCATCCATCCACTGTGATTTTCTTGCACGACGCTGC 1046
QY 150 GlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThrGlyVal 169
DB 1047 AATGTTGGATCTCGTACTCT----- 1067
QY 170 ValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeu 189
DB 1068 -----TCAAGTCT 1076
QY 190 IleAspThrThrAsnLeuAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSer 209
DB 1077 GTTCTAACACGCTCGCAGTGTGAAGACGCTATGCTTCTTGGAGTTGTTCTACCAG 1136
QY 210 GlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeuThrThrGluSer 229
DB 1137 CAATTCCCACTTG-----CTGAAGAACCACTTCCCATTCGCGCGGAGTCTG 1184
QY 230 TyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgile 249
DB 1185 TACGCTGTATTAATCCAGTGTGCTCCGAGATTCTCACCACATGCTGACAGA--- 1241
QY 250 AlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGly 269
DB 1242 -----TCTTCAACCTCCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1277
QY 270 IleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyr 289

```

```

DB 1278 TTGACCGACCCACTTAACAGTACCCATCTACGAGAGATGGCATGCTCTACTGATGGT 1337
QY 290 GlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetPro 309
DB 1338 GGCTATGAGCAA---CCCTGACGAGTCTGAGTCCGAAGGAATGTTTGGAGACCTTGGCT 1394
QY 310 AsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAsp 329
DB 1395 AGA---TGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1436
QY 330 TyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyr 349
DB 1437 GTGTTCTCATGTCCTCCGCTCCATCTACTGCAACACGACCACTTGGACCATTCACAA 1496
QY 350 AlaPheAlaGlyArgGlyValTyrAspIleArgHisPro-----Tyr 363
DB 1497 AAG---ACCGGCGAAGACGCTGACGACGTTAGAAAGATGTGCGAGGGAACCTGTGCTG 1553
QY 364 AspAspProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 383
DB 1554 AAAGAC-----ATGGAATACATGACCAATATTTGAACACGACACTTTGTCAAGGAA 1604
QY 384 AlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsnAspValTyrTyrAlaPhe 402
DB 1605 AAGTTGGCGGTGAGTTGACACTTACGAGTCTGTAATTTCCGACGTGAACAGAACTTC 1664
QY 403 GlnGlnThrGlyAspPheValTrpPro-----AsnPheIleGluAspLeuGlu 419
DB 1665 CTGTTTGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1724
QY 420 IleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAlaAspTyrIleCys 438
DB 1725 -----GGTCTTCTCTGTC-----CTGATTTACGAGGAGACAGAGTTTCATCTGC 1769
QY 439 AsnTrpPheGlyGlyGluAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe 458
DB 1770 AATTGGCTCGGAACCAACAGCTGTTCCATGAGTCCCTTGTCTGGACACGATGAATTC 1829
QY 459 ArgSerAlaGly---TyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArg 477
DB 1830 GAGTCCCCGAGCTGTACACCTCACCCTTGAAGGATGGCACTAAGTCCGCGAGGTCAG 1889
QY 478 GluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyr 497
DB 1890 AATGCTGGCAAGTTCACCTTCTGCTAGATGTTTGTGAGGAGACACATGTTCCATAGC 1949
QY 498 GlnProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGlu 517
DB 1950 CAGCCTGAGAGCTCTTGTGCTATGTCATAGATGATGATGATGATGATGATGATGATG 2006
QY 518 GlyGlnLysLys 521
DB 2007 GGAACCAAGAAA 2018

RESULT 2
US-09-938-842A-683
; Sequence 683, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: K-reps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16

```

```

; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 683
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-683

```

Alignment Scores:

Pred. No.:	1,09e-45	Length:	1551
Score:	475.00	Matches:	152
Percent Similarity:	42.72%	Conservative:	68
Best Local Similarity:	29.51%	Mismatches:	189
Query Match:	15.94%	Indels:	106
DB:	10	Gaps:	20

US-09-712-338-2 (1-555) x US-09-938-842A-683 (1-1551)

10 LeuProLeuValAlaAlaSerTrpAlaLeuProGlySerThrProAlaSerValGlyArg 23
196 TTGCTCTCTCACTGCTGCG-----GAGGACCGGGGATTGTGTAACGG 237

30 ArgGlnLeuProLysAsn-----ProThrGlyValLysThrLeuThrAlaAsn 47
238 AAGTTCGTATCCCGAATATATCTGCCGATGGTGGCCCTACTGFCGACGATTTAGTGCAT 297

48 ValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyVal 67
298 CATGCTGGTTATTACAAGCTCCGAAATCTCGTGGC----- 333

68 LysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheTrpPhePhe 87
334 -----GCAAGCATGTTCTACTTCTTCTTCTTC 357

88 GluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGly 107
358 GAGTCACCG--AACAAAGAGATGCTCTCTTGTGATTTGGTTGACGGAGGCGCTCGA 414

108 SerAspSerLeuIleGlyPheGluGluLeuGlyProCysHisValAsnSerThrPhe 127
415 TGTAGTAGTGAGTGGCTGTGTTCTAGAGAAATGGTCTTTCAAGATCACTAGTAACATG 474

128 AspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGln 147
475 TCTCTGTGTTGGATGAGTATGGATGGGATCAGCTTTCCAATCTTGTATGTTGATCAG 534

148 ProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThr 167
535 CTGTGTGAACCTGGTTTCAGTACACGACATAAAGTGCATATCCGTCATGACGAAATC 594

168 GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAla 187
595 GGAGTT----- 600

188 ThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGluIleLeuGlnGlyPhe 207
601 -----AGCAATCATCTTTATGATTTTCTGACGGCTTTC 633

208 LeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeuTrpThr 227
634 TTGCGGAGCACCCCTAAGTTG-----GCAAAAAACACATTTATATTACTGGA 681

228 GluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnGlu 247
682 GAGTCATATCTGGGCACTATATCCAGCTTTTCTCCCGAGTCCATTAAGGAACAAG 741

248 ArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIle 267
742 -----GCTAAT-----GAGGAGTTCATATTAACTTAAGGATTCGCTATTGGA 786

268 AsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaAsnAsn 287

RESULT 3

```

RESOLI 3
US-09-971-429B-33
; Sequence 33, Application US/09971429B
; Publication No. US20030175704A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K. W.
; APPLICANT: Shyjan, Andrew W.
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0040 US
; CURRENT APPLICATION NUMBER: US/09/971,429B
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/239,024
; PRIOR FILING DATE: 2000-04-10
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 33
; LENGTH: 1949
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175704A1 239680.4

```

FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: 71, 73
 ; OTHER INFORMATION: a, t, c, g, or other
 US-09-971-429B-33

Alignment Scores:

Pred. No.: 1.13e-32 Length: 1949
 Score: 366.50 Matches: 146
 Percent Similarity: 38.41% Conservative: 71
 Best Local Similarity: 25.84% Mismatches: 213
 Query Match: 12.30% Indels: 135
 DB: 12 Gaps: 25

US-09-712-338-2 (1-555) x US-09-971-429B-33 (1-1949)

QY 3 GlyTyrGluPheLeuSerValLeuPro-----LeuValAlaAlaSerTyr----- 17
 DB 338 GGAGCCCATTTCTTAACATAATACCTGAGAGATGGTTGGTGCCTATGGAGGTGATT 397
 QY 18 -----AlaLeuProGly-----SerThrProAla 25
 DB 398 GTTTCGTGCTGCTGTGTGATGCTGGCCCTCTGATGGCTCTTCACTCCCTATACAGA 457
 QY 26 SerValGlyArgGlnLeuProLysAsnProThrGlyValLysThrLeuThrAla 45
 DB 458 AGTGTTCCTCC-----ATGCCACTAAGGAGACTCAGGACGCCAATTAFTTCTCACCCCT 511
 QY 46 AsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrPro 65
 DB 512 TACATGAAGCTGGGAGATCCAAAGAGAGAAATTCAGTTGGTGGCCCTTCCCA 571
 QY 66 Gly-----ValSerTyrSerGlyTyr-----ValAspThrSerProGluSerHis 81
 DB 572 GGAATGAAATGAAGAGTTATGCGGCTTCCTCACCCTGAATAAGACTTACAAACAGCAAC 631
 QY 82 ThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTyr 101
 DB 632 CTCCTCTCTGCTCTCTCCAGCTCAGATACAGCAGAGATGCCCCAGTAGTCTCTCTGG 691
 QY 102 LeuAsnGlyGlyProGlySerAspSerLeuLeuGlyLeuPheGluGluLeuGlyProCys 121
 DB 692 CTACAGGTGGCGGAGGAGTTCATCCATGTTGGACTCTTTGTGAAACATGGGCTTAT 751
 QY 122 HisValAsnSerThrPheAspAspTyrIleAsnProHisSerTyrAsnGluValSerAsn 141
 DB 752 GTTGTACAAAGTAACATGACCTTCGCTGACAGAGACTTCCCTGGACCAACAGCTCTCC 811
 QY 142 LeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGly 161
 DB 812 ATGCTTTATCATTAACATCAAGTGGGACAGGCTTCAGTTTACTGATGATACCCACGGA 871
 QY 162 SerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArg 181
 DB 871 ----- 871
 QY 182 TyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTyr 201
 DB 872 -----TATGCAGTCAATGAGGAGGATGATGACGCGGATTATATACAGT 913
 QY 202 GluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLys 221
 DB 914 GCNCTAATTCAGTTTTTC-----CAGATATTTCTCCTGAATATAAATAAT 958
 QY 222 AspPheSerLeuThrPheGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHis 241
 DB 959 GACTTTTATGCTCACTGGGAGTCTTATGAGGGAATAATGTCGACCATTCGACACCTC 1018
 QY 242 PheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPhe 261
 DB 1019 ATCCATTCCTCAAC-----CCTGTGAGAGAGGTGAAGATCAACCTG 1060
 QY 262 AsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrPro 291

DB 1061 AACGAATTGCTATTGGAGATGGATATTTCTGATCCCGAATCAATATATAGGGGCTATGCA 1120
 QY 282 GluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyr 301
 DB 1121 GAATTC-----CTGTACCAATTTGGCTTGTGGATGAGAGCAAAAAAGTAC 1168
 QY 302 MetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGln 321
 DB 1169 -----TTCGAGAGCAG-----TGCCATGAATGATAGACACATCAGGAG 1210
 QY 322 ThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 338
 DB 1211 CAGAACTGTTTGAGGCTTTGAATACTGGATAAACTAGATAGGCGGACTTACCAAGT 1270
 QY 339 -----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGlyArg 354
 DB 1271 GATCCTTCTTACTTCCAGATGTTACAGAGATGTAGTAATTACTATACTTTTG----- 1324
 QY 355 GlyValTyrAspIleArgHisProTyrAspAspProThrProSerTyrTyrAsnLys 374
 DB 1325 -----CGGTGCACGGAACCTGAGGATCAGCTTTACTATGTGAAA 1363
 QY 375 PheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSer 394
 DB 1364 TTTTGTCTCTCCAGAGGTGAGACAGCCATCCAGTGGG-----AATCAGACTTTTAA 1420
 QY 395 AsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrProAsn 412
 DB 1421 GATGGAACATATAGTTGAAAAAGTACTTGCAGAGAGATACAGTACAGTCAGTAAAGCCATGG 1480
 QY 413 PheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyrGly 432
 DB 1481 TTAACGTGAATCATGAATAAT-----TATAAGTTCTGTGATCAATGCG 1525
 QY 433 AspAlaAspTyrIle-----CysAsnTyr 440
 DB 1526 CAACTGGACATCATCTGTCGACCTGCCCTGACAGAGCGCTCTTGTATGGGCGATGGACTGG 1585
 QY 441 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 457
 DB 1586 AAAGGATCCAG-----GAATACAAGAGGCGCAAAAAAGTTTGG 1627
 QY 458 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGlyVal 471
 DB 1628 AAGATCTTTAAATCTGACAGTGAAGTGGCTGTTAC----- 1663
 QY 472 GluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGly 491
 DB 1664 -----ATCGGCAAGCGGTGACTTCCATCAGGTAATATTTCGAGGTGGAGGA 1711
 QY 492 HisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIlePhe 511
 DB 1712 CATATTTTACCCTATGACCAGGCTCTGAGAGCTTTTGACATGATTAAATCGATTAT 1771
 QY 512 -----GlyTyrAsp 514
 DB 1772 GGAAGAGGATGGAT 1786

RESULT 4

US-10-084-018-4
 ; Sequence 4, Application US/10084018
 ; Publication No. US20020160499A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; Hawkins, Phillip R.
 ; Hillman, Jennifer L.
 ; Lal, Preeti
 ; Goli, Surya K.
 ; TITLE OF INVENTION: NOVEL HUMAN SERINE
 ; CARBOXYPEPTIDASE
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS: ;

ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/084.018
 FILING DATE: 25-Feb-2002
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/299.689A
 FILING DATE: <Unknown>
 APPLICATION NUMBER: 08/828.488
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0241 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1551 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: MMLR3D101
 CLONE: 566993
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-10-084-018-4

Alignment Scores:
 Pred. No.: 3 54e-32 Length: 1551
 Score: 361.00 Matches: 130
 Percent Similarity: 39.51% Conservative: 62
 Best Local Similarity: 26.75% Mismatches: 178
 Query Match: 12.12% Indels: 116
 DB: 13 Gaps: 21
 US-09-712-338-2 (1-555) x US-10-084-018-4 (1-1551)

QY	65	ProGly-----VallySerTyrSerGlyTyr-----ValAspThrSerProGluSer	80
DB	281	CCAGGACTGAACATGAAGATGTTATGCGGCTTCCTCAGCGTGAATAAGACTTACACAGC	340
QY	81	HisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu	100
DB	341	AACTCTTCTTGGTCTTCTCCAGCTCAGATACAGCCAGGATGCCCGAGTAGTTCTC	400
QY	101	TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro	120
DB	401	TGGCTACAGGTGGCGGGAGGTTTCATCCATGTTGGAGCTCTTTGGTGAACATGGGCT	460
QY	121	CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer	140
DB	461	TATGTTGTGCACAGTAACATGACCTTGGCTGACAGAGACTTCCCTGGACCAACGCTC	520
QY	141	AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp	160
DB	521	TCACTCTTACATTCATCAATCAGTGGGACAGGCTTCAGTTTACTGATGATACCCAC	580
QY	161	GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly	180
DB	581	GGA-----	583
QY	181	ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAla	200

DB	584	-----TATGCAATCAATGAGCAGCATGTAGCAGGATTTATAC	622
QY	201	TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer	220
DB	623	AGTGCACATAATTCAGTTTTC-----CAGATATTTCTCGAATATAAAAT	667
QY	221	LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn	240
DB	668	AATGACTTTTATGTCACTGGGAGCTTATGACGAGAAATATGTGCCAGCCATTCACAC	727
QY	241	HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn	260
DB	728	CTCATCCATTCCTCAAC-----CCTGTGAGAGAGGTGAGATCAAC	769
QY	261	PheAsnSerLeuGlyIleIleAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTyr	280
DB	770	CTGAACGGAATTGCTATTGGAGATGGATATCTGATCCCGAATCAATATAGGGGGCTAT	829
QY	281	ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn	300
DB	830	GCAGAATTC-----CTGTACCAAAATGGCTTGTGGATGAGAACAAAAAG	877
QY	301	TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys	320
DB	878	TAC-----TTCAGAGACAG-----TGCATGAATGCCATAGAACATCAGG	919
QY	321	GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr	337
DB	920	AAGCAGAACTGGTTTGAGGCTTTGAAATACTGGATAAACTACTAGTAGCGGACTTAACA	979
QY	338	Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly	353
DB	980	AGTGATCCTTCTTACTTCCAGAATGTTACAGGATGTAGTAATTAATACTTATTTTG	1036
QY	354	ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn	373
DB	1037	-----CGGTGCACGGAACCTGAGGATCAGCTTTACTATGTG	1072
QY	374	LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln	393
DB	1073	AAATTTTGTCACTCCAGAGGTGACACAGCCATCCACGTGGG-----AATCAGACTTTT	1129
QY	394	SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro	411
DB	1130	AATGATGGAACATATAGTTGAAAAAGTACTTCCGAGAGAGATACACTCAGTCAGTTAAGCCA	1189
QY	412	AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr	431
DB	1190	TGGTTAACTGAATATCAATAAT-----TATAAGGTTCTGATCTACAAT	1234
QY	432	GlyAspAlaAspTyrIle-----CysAsn	439
DB	1235	GGCCAACTGGACATCATCTGGCAGCTGCCCTCAGACAGCGCTCTTGTATGGCATGGAC	1294
QY	440	TrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaIleGln-----	457
DB	1295	TGAAAGGTTCCAG-----GAATACAAAGAGGACAGAAAAAAGTT	1336
QY	458	-----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly	470
DB	1337	TGGAAGATCTTTAAATCTCAGAGTGAAGTGGCTGGTTAC-----	1375
QY	471	ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla	490
DB	1376	-----ATCCGGCAAGTGGGTGACTCCATCAGGTAATTAATTCAGGTGGA	1420
QY	491	GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle	510
DB	1421	GGACATATTTTACCCTATGACCAGCCTCTGAGAGCTTTTGACATGATTAAATCGATTCA	1480
QY	511	Phe-----GlyTrpAsp	514

Db 1481 TATGGAAGAAGATGGCAT 1498

RESULT 5
US-09-796-753-39
; Sequence 39, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 39
; LENGTH: 1697
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)...(1528)
US-09-796-753-39

Alignment Scores:
Pred. No.: 7.05e-32 Length: 1697
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.05% Indels: 116
DB: 11 Gaps: 21
US-09-712-338-2 (1-555) x US-09-796-753-39 (1-1697)
QY 65 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 80
Db 299 CAGGACGTAACATGAGAGATGATCCGGCTTCCTCACCAGGAAATAGACTTACACAGC 358
QY 81 HisThrPhePheTrpPhePheGluAlaArgHisAsnProGluThrAlaProleThrLeu 100
Db 359 AACCTCTTCTTCTGCTTCTTCCAGCTCAGATACAGCCAGAGATGCCCGAGTAGTCTC 418
QY 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuLeuGlyLeuPheGluGluLeuGlyPro 120
Db 419 TGGCTACAGGGTGGCGCGGAGGTTCATCCATGTTGGACTCTTTGTGGAACATGGCCT 478
QY 121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTyrAsnGluValSer 140
Db 479 TATGTTGTCCAAAGTAACATGACCTTGGCTGACAGAGACTTCCCTGGACCAACAGCTC 538
QY 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
Db 539 TCCATGCTTTACATTGACAAATCCAGTGGGCACAGGCTTCAGTTTACTGATGATACCCAC 598
QY 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
Db 599 GGA----- 601
QY 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAla 200
Db 602 -----TATGCAGTCAATGAGGACGATGATAGCAGGATTTATATAC 640
QY 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
Db 641 AGTGCACATAATTCAGTCTTTC-----CAGATATTTCTCCTGAATATAAAAAAT 695
QY 221 LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProIlePheAsn 240
Db 696 AATGACTTTTATGTCATCTGGGAGCTTATGAGGAAATATGTGCCGACCATGTCACAC 745
QY 241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
Db 746 CTCATCCATCCCTCCAC-----CCTGTGAGAGAGGTGAAGATCAAC 787
QY 261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 280
Db 788 CTGACACGAATTGCTATTGGAGATGGATATTTCTGATCCCGCAATCAATTATAGGGGCTAT 847
QY 281 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 300
Db 848 GCAGAATTC-----CTGTACCAATTTGGCTTGTGGATGAGAGCAAAAAAAG 895
QY 301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
Db 896 TAC-----TTCCAGAGACAG-----TGCATGATGATGATAGAACATCAGG 937
QY 321 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 337
Db 938 AAGCAGAACTGGTTTGAGGCCCTTTGAAATACTGATAAACTAGATGCGGACTTAACA 997
QY 338 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 353
Db 998 AGTGATCCTTCTTACTTCCAGAAATGTACAGGATGTAGTAATTAATACTTATCTTTTG- 1054
QY 354 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 373


```

Db 1055 -----CGTGACGGAACCTGAGCATCAGCTTTACTATGTG 1090
Qy 374 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 393
Db 1091 AAATTTTGTCACTCCAGAGGTGAGACAAGCCATCCAGTGGG---AATCAGACTTTT 1147
Qy 394 SerAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 411
Db 1148 AATGATGGAACATATAGTTGAAAGTACTTGGCAGAAGATACAGTACAGTCAGTTAAGCCA 1207
Qy 412 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 431
Db 1208 TGGTTAACTGAATCATCAATAAT-----TATAAGGTTCTGATCTACAAT 1252
Qy 432 GlyAspAlaAspTyrIle-----CysAsn 439
Db 1253 GGCCAATGGACATCATCGTGGCAGCTGCCCTGACAGAGCGCTCCTGTATGGGGCATGGAC 1312
Qy 440 TrpPheGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGln----- 457
Db 1313 TGGAAAGGATCCAG-----GAATACAAGAGGCGAGAAAGTT 1354
Qy 458 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 470
Db 1355 TGAAGATCTTAAATCTGACAGTGAAGTGGCTGTAC----- 1393
Qy 471 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
Db 1394 -----ATCGCGAAGCGGGTGCATCTCCATCAGGTAATATTCGAGGTGGA 1438
Qy 491 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 510
Db 1439 GGACATATTTTACCCTATGACCAAGCTCTGAGAGCTTTTCACATGATTAATGATTCATT 1498
Qy 511 Phe-----GlyTrpAsp 514
Db 1499 TATGGAAGATGGGAT 1516

```

RESULT 6

US-09-909-320-163

Sequence 163, Application US/09909320

Patent No. US20020132240A1

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, A.

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth, J.

APPLICANT: Kijavlin, Ivar J.

APPLICANT: Mather, Jennie P.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

FILE OF INVENTION: Acids Encoding the Same

FILE REFERENCE: 10466-14

CURRENT APPLICATION NUMBER: US/09/909,320

CURRENT FILING DATE: 2002-01-04

```

; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-909-320-163

```

Alignment Scores:

Pred. No.:	9,71e-32	Length:	2076
Score:	359.00	Matches:	130
Percent Similarity:	39.51%	Conservative:	62
Best Local Similarity:	26.75%	Mismatches:	178
Query Match:	12.05%	Indels:	116
DB:	10	Gaps:	21

US-09-712-338-2 (1-555) x US-09-909-320-163 (1-2076)

Qy 65 ProGly-----ValIysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 80

Db 295 CCAGGACTCAACATGAAGAGTTATCCCGCTTCCTCCCGTGAATAAGACTTTACAACAGC 354

Qy 81 HisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100

Db 355 AACCTCTTCTTCTGTTCTTCCAGCTCAGATACAGCCAGAGATGCCCCAGTAGTTCTC 414

Qy 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 120

Db 415 TGGCTACAGGTGGCGCGGAGGTTTCATCCATGTTTGACTCTTTGTGGAACATGGGCT 474

Qy 121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 140

Db 475 TATGTTGTCAAGTACATACATACCTTGGTGGAGAGACTTCCCTTGACCAACACGCTC 534

Qy 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160

Db 535 TCCATGCTTTTACATTGACATCCAGTGGCAGAGCTTCAGTTTACTGATGATACCAC 594

Qy 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180

Mon Nov 24 13:41:13 2003

us-09-712-338-2.rnpb

595 GGA----- 597
181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 200
598 -----TATGCACTCAATGAGGAGGATGAGCAGGGATTATAC 636
201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
637 AGTGCACCTAATTCAGTTTTC-----CAGATATTTCTGGAATATAAAAT 681
221 LysAspPheSerLeuThrGlnSerTyrGlyGlyHisTyrGlyProAlaPhePheAsn 240
682 AATGACTTTTATGTCAGGGAGTCTTATCAGGGAATATGTCAGCCATTCGACAC 741
241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
742 CTCATCCATTCCTCAAC-----CCTGTGAGAGGTTGAAGATCAAC 783
261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 280
784 CTGAACGGAATTGCTATTGAGATGATATCTGATCCGATCAATTAATATAGGGGGCTAT 843
281 ProGluPheAlaValAsnAsnThrTyrGlyIleIleIleAlaValAsnGluThrValTyrAsn 300
844 GCAGAAATTC-----CTGTACCAGAAATGCTGTTGGATGAGAGCAAGCAAAAAAG 891
301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
892 TAC-----TTCAGAGAGCAG-----TGCCATGAATGCATAGACACATCAGG 933
321 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 337
934 AAGCAGAACTGCTTTGAGGCTTTGAATACTGGATAACTACTAGTGGCGACTTAACA 993
338 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 353
994 AGTGATCTCTTACTTCAGAAATGTTACAGATGTAGTAATTAATTAATCTTTTG--- 1050
354 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 373
1051 -----CGTGCACGGAACCTGAGGATCAGCTTACTATGTG 1086
374 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 393
1087 AAATTTTGTCACTCCAGAGGTGAGACAGCCATCCAGTGGGG---AATCAGACTTTT 1143
394 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro 411
1144 AATGATGGAACTATAGTTGAAAGTACTTCCGAGAGATACAGTACAGTACAGTAAAGCCA 1203
412 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 431
1204 TGCTTAAGTAACTAATCATGATATAT-----TATAAGGTTCTGATCTACAAT 1248
432 GlyAspAlaAspTyrIle-----CysAsn 439
1249 GGCCAACTGACATCATCGTGGCAGCTGCCCTGACAGAGCGCTCCTTGATGGCATGGAC 1308
440 TrpPheGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGln----- 457
1309 TGGAAAGGATCCAG-----GAATACAGAGGAGGAGAAAAAGTT 1350
458 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 470
1351 TGGAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1389
471 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
1390 -----ATCCCGCAAGCGGTGACTTCCATCAGTAATATTCAGAGTGA 1434
491 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 510
1435 GGACATATTTTACCTATGACCAGCCTTGAGAGCTTTTGACATGATTATTCGATTTCATT 1494

Qy 511 Phe-----GlyTyrAsp 514
Db 1495 TATGAAAAAGATGGAT 1512
RESULT 7
US-09-909-088B-163
; Sequence 163, Application US/09909088B
; Patent No. US20020146709A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Borstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,088B
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219

APPLICANT: Williams, P. Mickey
 APPLICANT: Wood, William, I.
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 TITLE OF INVENTION: Acids Encoding the Same

FILE REFERENCE: 10466-14
 CURRENT APPLICATION NUMBER: US/09/905,291A
 CURRENT FILING DATE: 2001-07-12
 PRIOR APPLICATION NUMBER: PCT/US00/04414
 PRIOR FILING DATE: 2000-02-22
 PRIOR APPLICATION NUMBER: US 60/143,048
 PRIOR FILING DATE: 1999-07-07
 PRIOR APPLICATION NUMBER: US 60/145,698
 PRIOR FILING DATE: 1999-07-26
 PRIOR APPLICATION NUMBER: US 60/146,222
 PRIOR FILING DATE: 1999-07-28
 PRIOR APPLICATION NUMBER: PCT/US99/20594
 PRIOR FILING DATE: 1999-09-08
 PRIOR APPLICATION NUMBER: PCT/US99/20944
 PRIOR FILING DATE: 1999-09-13
 PRIOR APPLICATION NUMBER: PCT/US99/21090
 PRIOR FILING DATE: 1999-09-15
 PRIOR APPLICATION NUMBER: PCT/US99/21547
 PRIOR FILING DATE: 1999-09-15
 PRIOR APPLICATION NUMBER: PCT/US99/23089
 PRIOR FILING DATE: 1999-10-05
 PRIOR APPLICATION NUMBER: PCT/US99/28214
 PRIOR FILING DATE: 1999-11-29
 PRIOR APPLICATION NUMBER: PCT/US99/28313
 PRIOR FILING DATE: 1999-11-30
 PRIOR APPLICATION NUMBER: PCT/US99/28564
 PRIOR FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: PCT/US99/28565
 PRIOR FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: PCT/US99/30095
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: PCT/US99/30911
 PRIOR FILING DATE: 1999-12-20
 PRIOR APPLICATION NUMBER: PCT/US99/30999
 PRIOR FILING DATE: 1999-12-20
 PRIOR APPLICATION NUMBER: PCT/US00/00219
 PRIOR FILING DATE: 2000-01-05
 NUMBER OF SEQ ID NOS: 423
 SEQ ID NO 163
 LENGTH: 2076
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-905-291A-163

Alignment Scores:
 Pred. No.: 9,71e-32 Length: 2076
 Score: 359.00 Matches: 130
 Percent Similarity: 39.51% Conservatives: 62
 Best Local Similarity: 26.75% Mismatches: 178
 Query Match: 12.05% Indels: 116
 DB: 21 Gaps:

US-09-712-338-2 (1-555) x US-09-905-291A-163 (1-2076)

QY	65	ProGly	-----ValAspThrSerProGluSer	80
DB	295	CGAGGACTGAACATGAAGATTATGCCGGCTTCTACCGTGAATAGACTTACACAGC	354	
QY	81	HistThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu	100	
DB	355	AACCTCTCTTCGGTCTTCCAGCTCAGATACAGCCAGAGATGCCCGCCAGTTCTC	414	
QY	101	TrpLeuAsnGlyProGlySerAspSerLeuIleGlyLeuPheGluGluGlyPro	120	
DB	415	TGGCTACAGGTGGCGGGAGTTCATCCACTGTTGGACTCTTTGGAAATATGGGCCT	474	
QY	121	CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer	140	
DB	475	TATGTTGTCAAGTACATGACCTTGGCTGACAGAGACTTCCCTTGGACCAACAGCTC	534	
QY	141	AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp	160	
DB	535	TCCATGCTTTTACATCAATCCAGTGGCAGCGCTTCAGTTTACTGATGATACCCAC	594	
QY	161	GlySerIleAsnProValThrGlyValGluAsnSerSerPheAlaGlyValGlnGly	180	
DB	595	GGA-----	597	
QY	181	ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla	200	
DB	598	-----TATGCAGTCAATGAGGACGATGATGACCGGATTTATATAC	636	
QY	201	TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer	220	
DB	637	AGTGCACATAATTCAGTTTTTC-----CAGATATTTCTCCATATATAAAT	681	
QY	221	LysAspPheSerLeuTyrThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsn	240	
DB	682	AATGACTTTTATGTCACCTGGGAGTCTTATGCAGGAATATATGTCGAGCCATTGCAC	741	
QY	241	HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn	260	
DB	742	CTCATCCATTCCTCCAC-----CCTGTGAGAGAGTGAAGATCAAC	783	
QY	261	PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr	280	
DB	784	CTGAACGGAATTCGATTCGAGATGATTCGATCCGGAATCAATATATATATATATAT	843	
QY	281	ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn	300	
DB	844	GCAGAAATTC-----CTGTACCAATTCGCTTGTGATGAGAGCAAAAGAG	891	
QY	301	TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys	320	
DB	892	TAC-----TTCAGAGAGCAG-----TGCCATGAATGATAGACACATCAGC	933	
QY	321	GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr	337	
DB	934	AAGCAGAACTGGTTGAGGCCCTTGAATATCTGGATAACTACTAGTGGGACTTAAACA	993	
QY	338	Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly	353	
DB	994	AGTGATCCTTCTTACTTCCAGATGTTACAGGATGTAGTAATTAATTAATTTTGTG	1050	
QY	354	ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn	373	
DB	1051	-----CGGTGCAAGCACTGAGGATCAGCTTACTACTATGTG	1086	
QY	374	LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln	393	
DB	1087	AAATTTTGTCTCCAGAGGTGAGACAGCAATCCAGTGGG-----AATCAGACTTTT	1143	
QY	394	SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro	411	
DB	1144	AATGATGGAACATATAGTTGAAAAGTACTTGCAGAGAGATACAGTACAGTCAAGTAA	1203	
QY	412	AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr	431	
DB	1204	TGTTTAACTGAATCATGAATAT-----TATAAGGTTCTGATCTCAAT	1248	
QY	432	GlyAspAlaAspTyrIle-----CysAsn	439	
DB	1249	GGCCAACTGGACATCATCTGGGAGCTGCCCTGACAGAGCGCTTGTATGGGATGGAC	1308	
QY	440	TrpPheGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln-----	457	
DB	1309	TGGAAGGATCCAG-----GAATACAGAGGAGGAGAGAGAGAGTT	1350	
QY	458	-----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly	470	
DB	1351	TGGAAGATCTTTAAATCTGACGTGAAGTGGCTGTTAC-----	1389	

Qy 471 ValGluTyrGlyGluThrArgGluTyrGlyAsnPhSerPheThrArgValTyrGluAla 490
Db 1390 -----ATCGGCAAGCGGGTGACTTCCTCAGGTAAATTCGAGGTGA 1434
Qy 491 GlyHisGluValProTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 510
Db 1435 GGACATATTTACCTATGACAGCCTCTGAGAGCTTTTGACATGATTAATCATTCATT 1494
Qy 511 Phe-----GlyTyrAsp 514
Db 1495 TATGAAAGGATGGAT 1512

RESULT 9

US-09-902-853-163
; Sequence 163, Application US/09902853
; Publication No. US20020192659A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,853
; PRIOR FILING DATE: 2001-07-10
; PRIOR FILING DATE: 2000-09-18
; PRIOR FILING DATE: 1999-07-07
; PRIOR FILING DATE: 1999-07-26
; PRIOR FILING DATE: 1999-07-28
; PRIOR FILING DATE: 1999-09-08
; PRIOR FILING DATE: 1999-09-13
; PRIOR FILING DATE: 1999-09-15
; PRIOR FILING DATE: 1999-09-15
; PRIOR FILING DATE: 1999-09-15
; PRIOR FILING DATE: 1999-09-15
; PRIOR FILING DATE: 1999-10-05
; PRIOR FILING DATE: 1999-11-29
; PRIOR FILING DATE: 1999-11-30
; PRIOR FILING DATE: 1999-12-02
; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-902-853-163
Alignment Scores:
Pred. No.: 9.71e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.05% Indels: 116
DB: 10 Gaps: 21
US-09-712-338-2 (1-555) x US-09-902-853-163 (1-2076)
Qy 65 ProGly-----VallySerTyrSerGlyTyr-----ValAspThrSerProGluSer 80
Db 295 CCAGGACTGAACATGAAGAGTATCGCGGTCTCTCACCCTGATTAAGACTTACACAGC 354
Qy 81 HisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
Db 355 AACCTCTTCTCTGTTCTTCCAGCTCAGATACAGCCAGAGATGCCAGTAGTTCTC 414
Qy 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 120
Db 415 TGGCTACAGGTGGCGCGGAGTTCATCTCTTTGGAGCTCTTTGTGGAACATGGGCT 474
Qy 121 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSer 140
Db 475 TATGTTGTCAAGTAACATGACCTTGGGTGACAGAGACTTCCCTGGACCAACAGCTC 534
Qy 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
Db 535 TCCATGCTTTATCATTCACATCCAGTGGCAGAGCTTCAGTTTACTGATGATACCCAC 594
Qy 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
Db 595 GGA-----597
Qy 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAla 200
Db 598 -----TATGCAGTCAATGAGGACGATGACGAGCTAGCAGCGGATTATAC 636
Qy 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
Db 637 AGTGCACTAAITCAGTTTTC-----CAGATATTTCTCTGAATATAAATA 681
Qy 221 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn 240
Db 682 AATGACTTTTATGTCACCTGGGAGCTTATGACGAGAAATATGTCCGCAATTCACAC 741
Qy 241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
Db 742 CTCATCCATTCCTCAAC-----CCTGTGAGAGAGGTGAGATCAAC 783
Qy 261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 280
Db 784 CTGAACGGAATTTGCTATTTGGAGATGATATTCCTGATCCGCAATCAATATATAGGGGCTAT 843
Qy 281 ProGluPheAlaValAsnAsnThrTyrGlyIleIleValAlaValAsnGluThrValTyrAsn 300
Db 844 GCAGAAATC-----CTGTACCAATTCGCTTTGGATGAGAGCAAAAAAG 891

301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
Db TAC-----TTCAGAACGAG-----TGCCATGAATGCATAGAACACATCAGG 933
321 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 337
Db AAGCAGAACTGGTTTGGAGCCCTTGAATACTGGATAAATCTAGATGGCGACTTAACA 993
338 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 353
Db AGTGATCCTCTTACTTCCAGAACTTACAGATGTAGTAATTACTATACTTTTGG--- 1050
354 ArgGlyValTyrAspIleHisProTyrAspAspProThrProProSerTyrTyrAsn 373
Db -----CGGTGCACGGAACCTGAGGATCAGCTTACTATCTG 1086
374 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 393
Db AATGATGAGCACTATAGTTGAAAGTACTTCCAGAGATACACAGTACAGTCACTTAAGCCA 1203
412 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 431
Db TGGTTTAACGAAATCATGATAAT-----TATAAGGTTCTGATCTACAAT 1248
432 GlyAspAlaAspTyrIle-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro 411
Db -----CysAsn 439
440 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 457
Db TGGAAAGGATCCAG-----GAATACAAGAGCAGCAAAAGT 1350
458 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 470
Db TGGAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1389
471 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
Db -----ATCCGGCAAGCGGTGACTTCCATCAGAGTAAATTCAGAGTGA 1434
491 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 510
Db GGCAATATTTACCTTATGACCGAGCTTCTGAGAGCTTTGACATGATTAATCGATTCAAT 1494
511 Phe-----GlyTrpAsp 514
Db TATGGAAGAGGATGGGAT 1512

RESULT 10
US-09-907-824-163
; Sequence 163, Application US/09907824
; Publication No. US20020197671A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.

APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,824
PRIOR FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 163
LENGTH: 2076
TYPE: DNA
ORGANISM: Homo Sapien
US-09-907-824-163

Alignment Scores:
Pred. No.: 9,71e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.05% Indels: 116
DB: 10 Gaps: 21

US-09-712-338-2 (1-555) x US-09-907-824-163 (1-2076)

Qy 65 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 80
Db 295 CCAGACTGACATGAAGAGTTATCGCGCTTCTCCACCGTGAATAAGACTTACACAGC 354
Qy 81 HisthrPhePheTrpPhePheGluAlaArgHisAsnProGluThrAlaProIleThrIleu 100

355 AACCTCTCTCTGGTCTTCCACAGCTCAGATACAGCCAGAGATGCCAGTAGTTCTC 414
QY
101 TTPLeuAsnGlyGlyProGlySerAspSerLeuLeuGlyLeuGlyPro 120
Db
415 TGCTACAGGTGGCGGGAGGTTCATCTGATTTTGGACTCTTTGGAAACATGGGCCT 474
QY
121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 140
Db
475 TATGTTCTGCACAGTAACATGATGACCTTCGCTGCACAGACATCCCTCGACCAACCCCTC 534
QY
141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
Db
535 TCCATGCTTTTACATGACATCAATCAGTGGGACAGCTTCAGTTTCTGATGATACCCAC 594
QY
161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
Db
595 GGA----- 597
QY
181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla 200
Db
598 -----TATGAGTCAATGAGGACGATGATAGACCGGATTTATAC 636
QY
201 TTPGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
Db
637 AGTGCATTAATTCAGTTTTC-----CAGATATTTCTCGAATATAAAAT 681
QY
221 LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn 240
Db
682 AATGACTTTTATGCTACTGGGAGTCTTATGAGGGAATATGTGCCAGCCANTGCACAC 741
QY
241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
Db
742 CTCATCCATTCCTCAAC-----CCTGTGAGAGAGGTGAAGATCAAC 783
QY
261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 280
Db
784 CTGAACGGAATGCTATTGGAGATGATATCTGATCCCGAATCAATATATAGGGGGCTAT 843
QY
281 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 300
Db
844 GCAGATTC-----CTGTACCAATTTGGCTTGTGGATGAGAGCAACMAAAAG 891
QY
301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
Db
892 TAC-----TTCAGAGAGCAG-----TGCCATGATGTCATAGAACACATCAGG 933
QY
321 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 337
Db
934 AAGCAGAACTGGTTGAGGCCCTTTGAAATACTGGATAAACTACTAGATGGCGACTTAACA 993
QY
338 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 353
Db
994 AGTGATCCTTCTTACTTCCAGATGTTACAGATGTAGTAATTAATTAATTAATTTTG---- 1050
QY
354 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 373
Db
1051 -----CGGTGCACGGAACCTGAGATGAGCTTACTATGTG 1086
QY
374 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 393
Db
1087 AAATTTTGTCTACTCCAGAGGTGAGACAAAGCAATCCAGCTGGGG-----AATCAGACTTTT 1143
QY
394 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 411
Db
1144 AATGATGGAACATAGTTGAAAAGTACTTGCAGAGAGATACAGTCAGTCAGTTAAGCCA 1203
QY
412 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 431
Db
1204 TGGTTTAACCTGAATCATGAATAAT-----TATAAGTCTCTGATCTACAAAT 1248
QY
432 GlyAspAlaAspTyrIle-----CysAsn 439
Db
1249 GGCCAACTGGACATCATCTGTCAGCTGCCCTGACAGAGCGCTCTCTGATGGCGATGAC 1308

QY
440 TTPPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 457
Db
1309 TGGAAAGAGTCCCGAG-----GAATACAAGAGGCGAGAAAAAAGTTT 1350
QY
458 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 470
Db
1351 TGGAAAGATCTTTAAATCTGACGTGAAGTGGCTGGTTTAC----- 1389
QY
471 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
Db
1390 -----ATCCGGCAGCGGGTCACTTCATCAGTAATATTCGAGGTGGA 1434
QY
491 GlyHisGluValProTyrTyrGlnProIleAlaSerIleuGlnLeuPheAsnArgThrIle 510
Db
1435 GGACATATTTTACCTATGACCAAGCTCTGAGAGCTTTTGCATGATTATATCGATTCAAT 1494
QY
511 Phe-----GlyTrpAsp 514
Db
1495 TATGGAAGAGGATGGGAT 1512
RESULT 11
US-09-907-841-163
; Sequence 163, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,841
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089


```

; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-904-011-163

Alignment Scores:
Pred. No.: 9,71e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.05% Indels: 116
DB: 11 Gaps: 21

US-09-712-338-2 (1-555) x US-09-904-011-163 (1-2076)
QY 65 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 80
Db 295 CCAGGACTGAACATCAGAGGATGTCGGCTTCTCCACGGTAATAGACTTACACAGC 354
QY 81 HisThrPhePheTyrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
Db 355 AACCTCTTCTTCTGGTTCTTCCAGCTCAGATCAGCCAGAGATGCCCGAGTAGTTCTC 414
QY 101 TtpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 120

```

108 PRIOR FILING DATE: 1999-11-29
109 PRIOR APPLICATION NUMBER: PCT/US99/28313
110 PRIOR FILING DATE: 1999-11-30
111 PRIOR APPLICATION NUMBER: PCT/US99/28564
112 PRIOR FILING DATE: 1999-12-02
113 PRIOR APPLICATION NUMBER: PCT/US99/28565
114 PRIOR FILING DATE: 1999-12-02
115 PRIOR APPLICATION NUMBER: PCT/US99/30095
116 PRIOR FILING DATE: 1999-12-16
117 PRIOR APPLICATION NUMBER: PCT/US99/30911
118 PRIOR FILING DATE: 1999-12-20
119 PRIOR APPLICATION NUMBER: PCT/US99/30999
120 PRIOR FILING DATE: 1999-12-20
121 PRIOR APPLICATION NUMBER: PCT/US00/00219
122 PRIOR FILING DATE: 2000-01-05
123 NUMBER OF SEQ ID NOS: 423
124 SEQ ID NO 163
125 LENGTH: 2076
126 TYPE: DNA
127 ORGANISM: Homo Sapien
128 US-09-906-742-163
129
130 Alignment Scores:
131 Pred. No.: 9,71e-32 Length: 2076
132 Score: 359.00 Matches: 130
133 Percent Similarity: 39.51% Conservative: 62
134 Best Local Similarity: 26.75% Mismatches: 178
135 Query Match: 12.05% Indels: 116
136 DB: 11 Gaps: 21
137
138 US-09-712-338-2 (1-555) x US-09-906-742-163 (1-2076)
139 QY 65 Progly-----ValysSerTyrSerGlyTyr-----ValasThrSerProGluSer 80
140 Db 295 CCAGGACTGAACATGAAGAGTTATCCGGCTTCTCCACCGTGATTAAGACTTACACAGC 354
141 QY 81 HisThrPhePhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
142 Db 355 AACCTCTCTCTCTGGTCTTCCAGCTAGATACAGCCAGAGATGCCCGAGTAGTCTC 414
143 QY 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuPro 120
144 Db 415 TGGCTACAGGGTGGCGCGGAGGTTTCATCCATGTTTGGACTCTTTGTGGAACATGGCCT 474
145 QY 121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 140
146 Db 475 TATGTTGTCAACAGTAACATGACCTTGGCTGACAGAGACTTCCCTGGACCAACACGCTC 534
147 QY 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
148 Db 535 TCATGCTTTTACATTCAGCAATCCAGTGGGCACAGGCTTCAGTTTACTGATGATCCAC 594
149 QY 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
150 Db 595 GGA----- 597
151 QY 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaAla 200
152 Db 598 -----TATGAGTCAATGAGGACGATGATGACCGGATTTATAC 636
153 QY 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
154 Db 637 AGTGCACATATTCAGTTTTC-----CAGATATTTCTCTGAATATAAAAT 681
155 QY 221 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsn 240
156 Db 682 AATGACTTTTATCTCCTGGGAGTCTTATGAGGGAATATATGCGAGCATTCGCACAC 741
157 QY 241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
158 Db 742 CTATCCATTCCTCCCTCAAC-----CCTGTGAGAGAGGTGAAGATCAAC 783
159 QY 261 PheAsnSerLeuGlyIleIleAsnGlyIleAsnGlyIleAsnGlyIleAlaProTyrTyr 280

458 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 470
459 TGGAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1389
460 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
461 ATCCGCAAGCGGGTGACCTGACCTCCATCAGGTAATATTTCGAGGTGGA 1434
462 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 510
463 GGACATATTTTACCTATGACCGCTCTGAGAGCTTTTGACATGATTAATCGATCAAT 1494
464 511 Phe-----GlyTyrAsp 514
465 1495 TATGGAAGAGTGGAT 1512
466
467 RESULT 13
468 US-09-906-742-163
469 Sequence 163, Application US/09906742
470 Publication No. US20030023054A1
471 GENERAL INFORMATION:
472 APPLICANT: Genentech, Inc.
473 APPLICANT: Ashkenazi, Avi
474 APPLICANT: Botstein, David
475 APPLICANT: Desnoyers, Luc
476 APPLICANT: Eaton, Dan L.
477 APPLICANT: Ferrara, Napoleone
478 APPLICANT: Filvaroff, Ellen
479 APPLICANT: Fong, Sherman
480 APPLICANT: Gao, Wei-Qiang
481 APPLICANT: Gerber, Hanspeter
482 APPLICANT: Gertitsen, Mary E.
483 APPLICANT: Goddard, A.
484 APPLICANT: Godowski, Paul J.
485 APPLICANT: Grimaldi, Christopher J.
486 APPLICANT: Gurney, Austin L.
487 APPLICANT: Hillan, Kenneth, J.
488 APPLICANT: Kljavin, Ivar J.
489 APPLICANT: Mather, Jennie P.
490 APPLICANT: Pan, James
491 APPLICANT: Paoni, Nicholas F.
492 APPLICANT: Roy, Margaret Ann
493 APPLICANT: Stewart, Timothy A.
494 APPLICANT: Tumas, Daniel
495 APPLICANT: Williams, P. Mickey
496 APPLICANT: Wood, William, I.
497 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
498 FILE REFERENCE: 10466-14
499 CURRENT APPLICATION NUMBER: US/09/906,742
500 CURRENT FILING DATE: 2001-07-16
501 PRIOR APPLICATION NUMBER: 09/665,350
502 PRIOR FILING DATE: 2000-09-18
503 PRIOR APPLICATION NUMBER: PCT/US00/04414
504 PRIOR FILING DATE: 2000-02-22
505 PRIOR APPLICATION NUMBER: US 60/143,048
506 PRIOR FILING DATE: 1999-07-07
507 PRIOR APPLICATION NUMBER: US 60/145,698
508 PRIOR FILING DATE: 1999-07-26
509 PRIOR APPLICATION NUMBER: US 60/146,222
510 PRIOR FILING DATE: 1999-07-28
511 PRIOR APPLICATION NUMBER: PCT/US99/20594
512 PRIOR FILING DATE: 1999-09-08
513 PRIOR APPLICATION NUMBER: PCT/US99/20944
514 PRIOR FILING DATE: 1999-09-13
515 PRIOR APPLICATION NUMBER: PCT/US99/21090
516 PRIOR FILING DATE: 1999-09-15
517 PRIOR APPLICATION NUMBER: PCT/US99/21547
518 PRIOR FILING DATE: 1999-09-15
519 PRIOR APPLICATION NUMBER: PCT/US99/23089
520 PRIOR FILING DATE: 1999-10-05
521 PRIOR APPLICATION NUMBER: PCT/US99/28214

```

Db      784 CTGAACGAATGCTATTGGAGATGGATATCTCGATCCGCAATCAATTATAGGGGCTAT 843
QY      ProGluPheAlaValAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 300
Db      844 GCAGAAATTC-----CTGTACCAATTCGCTTGGATGAGAGCAAGCAAAAAG 891
QY      TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
Db      892 TAC-----TTCCAGAACAG-----TGCCATGAATGCATAGAACACATCAGG 933
QY      GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 337
Db      934 AAGCAGAACTGGTTGAGGCCTTTGAAATATCTGGATAAATACTAGATGGCGACTTACA 993
QY      Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 353
Db      994 AGTGATCTCTTACTTCCAGAAATGTTACAGATGTAGTAATTAATACTTTTGTG--- 1050
QY      ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProCserTyrTyrAsn 373
Db      1051 -----CGGTGCACGGAACCTGAGGATCAGCTTACTATGTG 1086
QY      LysPheLeuAlaLysAspServAlMetAspAlaIleGlyValAsnIleAsnTyrThrGln 393
Db      1087 AATTTTGTCACTCCAGAGGTGAGACAGCCATCCACGTGGGG---AATCAGACTTTT 1143
QY      SerAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 411
Db      1144 AATGATGAACTATAGTTAAAGTACTTCCGAGAAGATACAGTACGTAGTAAAGCCA 1203
QY      AsnPheIleGluAspLeuGluGluLeuAlaLeuProValArgValSerLeuIleTyr 431
Db      1204 TGTTAACTGAATCATCAATAAT-----TATAAGGTTCTGATCTACAAT 1248
QY      GlyAspAlaAspTyrIle-----CysAsn 439
Db      1249 GGCAACTGGACATATCTGGCAGCTGCCCTGACAGAGCGCTCCTTGATGGCATGGAC 1308
QY      TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 457
Db      1309 TGGAAAGGATCCAG-----GAATACAGAAAGGAGCAAAAAGTT 1350
QY      PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 470
Db      1351 TGAAGATCTTTAAATCTCAGAGTGAAGTGGCTGTAC----- 1389
QY      ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
Db      1390 -----ATCCGCAAGCGGTGACTTCCATCAGGTAATTTATCGAGGTGA 1434
QY      GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnPheAsnArgThrIle 510
Db      1435 GGACATATTTTACCTATGACCAAGCTCTGAGAGCTTTTGACATGATTAATCGATTCATT 1494
QY      Phe-----GlyTrpAsp 514
Db      1495 TATGGAAGAAAGATGGAT 1512

RESULT 14
US-09-906-838-163
; Sequence 163, Application US/09906838
; Publication NO. US20030027143A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang

```

```

; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Guney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,838
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-906-838-163

Alignment Scores:
Pred. No.: 9,71e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.05% Indels: 116
DB: 11 Gaps: 21

US-09-712-338-2 (1-555) x US-09-906-838-163 (1-2076)

```

QY 65 ProGly-----VallySerTyrSerGlyTyr-----ValAspThrSerProGluSer 80
 Db 295 CCAGACTGAACAGAGATTATGCGGCTTCCTCCCGTGAATAAGACTTACAAACAGC 354
 QY 81 HisThrPhePhePhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
 Db 355 AACCTCTCTCTCTCTCCAGCTCAGATACAGCCAGAGATGCCCGAGTAGTTCTC 414
 QY 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluGlyPro 120
 Db 415 TGGTACAGGTGGCGCGGAGTTCATCCAGTTTGGACCTTTTGTGCAATGGGCT 474
 QY 121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTyrAsnGluValSer 140
 Db 475 TATGTTGTACAGTAATACATGACCTTGGTGACAGAGACTTCCCGTGGACCAACAGCTC 534
 QY 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
 Db 535 TCCATGCTTACATTCACAATCCAGTGGGCACAGGCTTCAGTTTACTGATGATACCCAC 594
 QY 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
 Db 595 GGA----- 597
 QY 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 200
 Db 598 -----TATGAGTCAATGAGGAGGAGTGTAGTACCGGATTTATAT 636
 QY 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
 Db 637 AGTGCACTAATTCAGTTTTC-----CAGATATTTCTCGAATATATAAAT 681
 QY 221 LysAspPheSerLeuThrGluSerTyrGlyHisTyrGlyProAlaPheAsn 240
 Db 682 AATGACTTTTATGCTACTGGGAGTCTTATGCGGGAATATGTGCCACCCATTGCCACAC 741
 QY 241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
 Db 742 CTCATCCATTCCTCAAC-----CCTGTGAGAGAGGTGAAGATCAAC 783
 QY 261 PheAsnSerLeuGlyIleLeuAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTyr 280
 Db 784 CTGACCGGAATGCTTGTGAGATGATATCTGATCCCGAATCAATATATAGGGGCTAT 843
 QY 281 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 300
 Db 844 GCAGAAATC-----CTGTACCAATTTGGCTTGTGATGAGAAGCAAAAAAG 891
 QY 301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
 Db 892 TAC-----TTCCAGAGAGCAG-----TGCCATGAATGCATAGAACACATCAGG 933
 QY 321 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 337
 Db 934 AAGCAGAACTGGTTTGAAGGCTTTGAATACTCGATAAATACTAGATGCGGACTTAACA 993
 QY 338 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 353
 Db 994 AGTGATCCTCTTACTTCCAGATGTTACAGGATGTAGTAATATACATACTTTTG--- 1050
 QY 354 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 373
 Db 1051 -----CGGTGCAAGCAACCTGAGGATCAGCTTACTATGTG 1096
 QY 374 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 393
 Db 1087 AAATTTTGTCACTCCAGAGGTGAGCAACCCATCCACGGGGG-----AATCAGACTTTT 1143
 QY 394 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 411
 Db 1144 AATGATGGAATACTAGTTGAAAGTCTTTCGAGAGAGATACAGTACAGTCAAGTAAAGCCA 1203

QY 412 AsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr 431
 Db 1204 TGGTAACTGAATCAATAAAT-----TATAAGTTTCTGATCTACAAT 1248
 QY 432 GlyAspAlaAspTyrIle-----CysAsn 439
 Db 1249 GCCCACTGGACATCATCTGGCAGCTGCCCTGACAGAGCGCTCCTTGATGGCATGGAC 1308
 QY 440 TrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGln----- 457
 Db 1309 TGGAAAGGATCCAG-----GAATACAAAGAGCGCAGAAAAAGTT 1350
 QY 458 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 470
 Db 1351 TGGAGATCTTTAATCTGACAGTCAAGTGGCTGGTTAC----- 1389
 QY 471 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
 Db 1390 -----ATCCGCAAGCGGTGACTTCCATCAGGTAAATTATTCGAGGTGGA 1434
 QY 491 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 510
 Db 1435 GCACATATTTTACCTATGACACCGCTCTGAGAGCTTTTGACATGATTAATCGATTCATT 1494

QY 511 Phe-----GlyTyrAsp 514
 Db 1495 TATGAAAAGGATGGAT 1512

RESULT 15
 US-09-907-613-163
 ; Sequence 163, Application US/09907613
 ; Publication No. US20030027145A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, A.
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Hillan, Kenneth, J.
 ; APPLICANT: Kljavin, Ivar J.
 ; APPLICANT: Mather, Jennie P.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Roy, Margaret Ann
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Williams, P. Mickey
 ; APPLICANT: Wood, William, I.
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 ; FILE OF INVENTION: Acids Encoding the Same
 ; FILE REFERENCE: 10466-14
 ; CURRENT APPLICATION NUMBER: US/09/907,613
 ; CURRENT FILING DATE: 2001-07-17
 ; PRIOR APPLICATION NUMBER: PCT/US00/04414
 ; PRIOR FILING DATE: 2000-02-22
 ; PRIOR APPLICATION NUMBER: US 60/143,048
 ; PRIOR FILING DATE: 1999-07-07
 ; PRIOR APPLICATION NUMBER: US 60/145,698
 ; PRIOR FILING DATE: 1999-07-26
 ; PRIOR APPLICATION NUMBER: US 60/146,222
 ; PRIOR FILING DATE: 1999-07-28
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594
 ; PRIOR FILING DATE: 1999-09-08

```
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-907-613-163

Alignment Scores:
Pred. No.: 9,71e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.05% Indels: 116
DB: 11 Gaps: 21

US-09-712-338-2 (1-555) x US-09-907-613-163 (1-2076)
QY 65 ProGly-----VallySerThrSerGlyTyr-----ValAspThrSerProGluSer 80
Db 295 CCAGGACTGAACATGAAGAGTATGCGGCTTCTCCACCGTGAATGAAGACTTACAACAGC 354
QY 81 HisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
Db 355 AACCTCTTCTTCTGTTCTTCCAGCTCAGATACACCGCAGAGATGCCCGAGTGTCTC 414
QY 101 TrpLeuAsnGlyClyProGlySerAspSerLeuIleGlyLeuPheGluGluGlyPro 120
Db 415 TGGCTACAGGGTGGGCGGAGGTTTCATCCATGTTTGGACTCTTTGTGGAACATGGGCT 474
QY 121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 140
Db 475 TATGTTGTCACAGTAACATGACCTTGGCTGACAGAGACTTCCCTGGACCAACAGCTC 534
QY 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
Db 535 TCCATGCTTTACATTGACAAATCCAGTGGCAGGCTTCAGTTTACTGATGATACCCAC 594
QY 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
Db 595 GGA----- 597
QY 181 ArgTyrProThrIleAspAlaThrIleLeuIleAspThrThrAsnLeuAlaGluAla 200
Db 598 -----TATGAGTCAATGAGGACGATGATGACGCGGATTTATAC 636
QY 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
Db 637 AGTGACATTAATTCAGTTTTC-----CAGATATTCTCTGAATATAAAAT 681
```

```
QY 221 LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn 240
Db 682 AATGACTTTTATGTCACCTGGGAGTCTTATCGAGGAATATATGTCAGCATTGCACAC 741
QY 241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGluLeuAsn 260
Db 742 CTCATCCATTCCCTCAAC-----CCTGTGAGAGAGGTGAAGATCAAC 783
QY 261 PheAsnSerLeuGlyIleIleAsnGlyIleAspGluAlaIleGlnAlaProTyrTyr 280
Db 784 CTGAACGGAATGCTATTGAGATGATATCTTCATCCCAATCAATATATAGGGGCTAT 843
QY 281 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 300
Db 844 GCAGATTTC-----CTGTACCAAAATTTGGCTTGTGGATGAGAGCAAAAAAG 891
QY 301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
Db 892 TAC-----TTCCAGAAGCAG-----TGCCATGAATGCATAGAACACATCAGG 933
QY 321 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 337
Db 934 AAGCAGAACTGGTTTCAGGCGCTTTGAAATACTGATAAACTACTAGATGCGGACTTAACA 993
QY 338 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 353
Db 994 AGTGATCCTTCTTACTTTCAGATGTTTACAGGATGTAGTAATTAATTAATCTTTTGTG-- 1050
QY 354 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 373
Db 1051 -----CGTGCACGGAACCTGAGGATCAGCTTTACTATGTG 1086
QY 374 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 393
Db 1087 AAATTTTCTCTCCAGAGGTGACAGCAAGCCATCCACGTGGGG--AATCAGACTTTT 1143
QY 394 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro 411
Db 1144 AATGATGGAACATATAGTTGAAAAGTACTTGCAGAGAGATACAGTACAGTCACTTAAGCCA 1203
QY 412 AsnPheIleGluAspLeuGluGluLeuAlaLeuProValArgValSerLeuIleTyr 431
Db 1204 TGGTTAACTGAATCATGATTAAT-----TATAAGGTTCTGATCTCAAT 1248
QY 432 GlyAspAlaAspTyrIle-----CysAsn 439
Db 1249 GGCCAACTGGACATCATCTGTCAGCTGCCCTGACAGAGCGCTCCTTGATGGCATGGAC 1308
QY 440 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 457
Db 1309 TGAAGAAGATCCAG-----GAATACAGAAGAGCGCAAAAAAAGTT 1350
QY 458 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 470
Db 1351 TGAAGAATCTTTAAATCTGACATGCAAGTGGCTGTTAC----- 1389
QY 471 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 490
Db 1390 -----ATCCGGCAAGCGGTGACTTCCATCAGGTAATTAATTATTCGAGGTGGA 1434
QY 491 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 510
Db 1435 GGACATATTTTACCCTATGACCAGCCTCTGAGAGCTTTTGACATGATTAATCGATTCAAT 1494
QY 511 Phe-----GlyTrpAsp 514
Db 1495 TATGGAAAAGGATGGAT 1512
```

Search completed: November 21, 2003, 20:29:39
Job time : 508 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 18:05:23 ; Search time 4885 Seconds
(without alignments)
4647.870 Million cell updates/sec

Title: US-09-712-338-2

Perfect score: 2979

Sequence: 1 MRGYEFLSVLPVLAASWALP.....HTQSSVPLPTATMSSSVGMA 555

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/UGPTO.spool.p/US09712338/runat_17112003_170522_11533/app_query.fasta_1.711
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CGN 1 1 4958 @runat_17112003_170522_11533 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.on:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pla:*
35: em.htg.rod:*
36: em.htg.man:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2975	99.9	1662	6	AR129928 Sequence
2	2618	87.9	2245	8	AF394242 Aspergill
3	2376	79.8	1656	6	AX534871 Sequence
4	2021	67.8	3150	6	AX534814 Sequence
5	980.5	32.9	1872	6	AX534872 Sequence
6	977	32.8	3221	6	AX534815 Sequence
7	944.5	31.7	1665	6	AX534866 Sequence
8	925	31.1	1581	6	AX534824 Sequence
9	902.5	30.3	2940	6	AX534767 Sequence
10	893	30.0	3080	6	AX534809 Sequence
11	652.5	21.9	4308	8	AX534809 Sequence
12	652.5	21.9	37000	8	SPAC19G12
13	636	21.3	1527	6	AX594884
14	636	21.3	2016	6	AX594884
15	636	21.3	2027	6	SCBRI139W
16	636	21.3	29686	6	AX536454
17	636	21.3	50277	2	SC29711
18	630	21.1	1611	6	AC138524_5
19	623	20.9	2002	6	AX534826
20	623	20.9	2002	6	I33983
21	623	20.9	2002	6	I74375
22	596.5	20.0	2509	8	I77239
23	596	20.0	2509	8	AF085063
24	595.5	20.0	1985	8	YACARPEPY
25	594.5	20.0	2214	8	PAU67174
26	594.5	20.0	3850	6	E12103
27	594	20.0	3850	8	PPERCIGEN
28	594	19.9	2068	6	I33982
29	594	19.9	2068	6	I74374
30	580	19.5	2660	6	I77238
31	568.5	19.1	2503	6	AX534769
32	568.5	19.1	2503	6	A75535
33	548.5	18.4	2280	8	I70282
34	541.5	18.2	1653	6	AB051820
35	527.5	17.7	2632	8	AX536918
36	527.5	17.7	15581	8	YSCPRCPY
37	523.5	17.6	2632	6	SCB175
38	479	16.1	1787	8	ARI29959
39	476.5	16.0	1772	8	AY072822
40	475	15.9	1551	6	RICCBP31
41	475	15.9	1551	6	AX412745
42	475	15.9	1551	6	AX505988
43	475	15.9	1551	8	AY149954
44	473.5	15.9	1771	8	AY091767
45	470.5	15.8	1816	8	AF141384
			3298	8	ABGSCFPZ

ALIGNMENTS

RESULT 1

AR129928
LOCUS AR129928 1662 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6187578.
ACCESSION AR129928
VERSION AR129928.1 GI:14117825
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (Bases 1 to 1662)
Blinkovsky, A., Berka, R., Rey, M., Golightly, E., Klotz, A.,
Mathisen, T., Erik., Dammann, C. and Brown, K.M.
Carboxypeptidases and nucleic acids encoding the same
Patent: US 6187578-A 1 13-FEB-2001;
JOURNAL Location/Qualifiers
FEATURES
source
1..1662
/organism="unknown"
BASE COUNT 396 a 468 c 398 g 400 t
ORIGIN

Alignment Scores:
Pred. No.: 1,04e-221 Length: 1662
Score: 2375.00 Matches: 554
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.87% Indels: 0
DB: 6 Gaps: 0

US-09-712-338-2 (1-555) x AR129928 (1-1662)

QY 1 MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuPro 20
DB 1 ATGGGTGGCTACGAATTTCTTCAGTGTACCTTGGTTGCAGCCAGTGGGCCCTTCCA 60
QY 21 GlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGlyValLys 40
DB 61 GGAAGTACACCGCGCTCGGTAGAGACAGCTACCCAGAACCCACCGGGGTCAAG 120
QY 41 ThrLeuThrThraAlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal 60
DB 121 ACTTCTTACAAACCGCAACATGTACCATCCGGTACAAGAACCCCGGGGCGATC 180
QY 61 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer 80
DB 181 TCGAGAGTACCCCGGTGTCAATCCTACTCTGGATATGTGCACACCTCTCCCGAGTCC 240
QY 81 HisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
DB 241 CATACCTTCTCTGGTCTTCGAAGCCAGACATACCCAGAACTGCACCTATCACATTG 300
QY 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 120
DB 301 TGGTTGAAGTGGCCCTGGAAGGATCTTTGATCGGTCTCTTCGAAGAGTTGGGCCCT 360
QY 121 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerThrAsnGluValSer 140
DB 361 TGCATGTCAATTTCGATTTGATGACTACATCAACCTCCTCCTGCGAAGCGAGTCTCC 420
QY 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
DB 421 AATTTACTATCTCTCCAGCCATTTGGAGTGGGTCTTTTCATATAGTATACGGTTGAT 480
QY 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
DB 481 GGGTCCATTAACCTGTAACTGGGGTGTGCGAAATTCGAGCTTTGCAGGAGTTCAGGGC 540
QY 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla 200
DB 541 CGGTACCCACCATTTGACCTCTCTGATGATACCTACCAATCTTGCAGGAGGGCGCT 600
QY 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
DB 601 TGGGAGATCTTCGAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTGCAGTCT 660

QY 221 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn 240
DB 661 AAGGACTTCAGTCTATGACGAGAGCTATGAGGGGCACTATGGTCTCTGCAATCTTCAAT 720
QY 241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
DB 721 CATTTTACGAGCAGCAATGAGAAATTCGCAACCGTAGTGTATATGGTTCAGCTTAAT 780
QY 261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 280
DB 781 TTCAACTCTCTGGGAATTTATTACCGCATCATCGACGAGCGATCGCCAGCCCTTACTAC 840
QY 281 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 300
DB 841 CCTGAATTCGCTGTGAACAATACCTACGTTATCAAGGCTGTCAAGACGCGCTCTACAC 900
QY 301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
DB 901 TACATGAAGTTTGCACCAATGCCAATGGTTGCCAGGATTTCGATTTCCACCTGCAAA 960
QY 321 GlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys 340
DB 961 CAGACAAACCGCACCGCATTAGCTGACTACGCCCTCTCGCCGAAAGCCACCAACATGTGC 1020
QY 341 ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 360
DB 1021 AGGGACAATGTTGAGGGGCCATACGCTTTGCTGTGTGGTGTGTATGATATTCGG 1080
QY 361 HisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSer 380
DB 1081 CATCCATATGATGACCGGACTCCGCAAGTATTACAAATTTCTGCAAGGACTCT 1140
QY 381 ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 400
DB 1141 GTCATGGACGCTATCGGCGTCAACATCAACTACACCCAGTCCCAATATGACCTCTACTAC 1200
QY 401 AlaPheGlnGlnThrGlyAspPheValTyrProAsnPheIleGluAspLeuGluIle 420
DB 1201 GCTTTCCACAAACAGGCGACTTTGCTGCGCCAACTTCATCGAAGCTTCGAGGAGATC 1260
QY 421 LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp 440
DB 1261 CTGTCTCTCCCGTGGGTGTCTCCCTCATCTATGGGACGCGGATACATCTCAACTGG 1320
QY 441 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 460
DB 1321 TTCGGCGGTGACGGCCGTTTCCCTCGCTGCGAACTACTCCCAAGCCGCCAGTTCGGAAGC 1380
QY 461 AlaGlyTyrThrProLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 480
DB 1381 GCAGGGTACACGCCCTCGAAGTCAACGGCGTGGATATGGGAAACTCGCGAGTATGGT 1440
QY 481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 500
DB 1441 AATTTCTCTCTACTCGGTCTATAGGACGCGCATGAAGTCCCATACTACACGCCCATC 1500
QY 501 AlaSerIleGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys 520
DB 1501 GCCTCCCTGCAATGTTTAAACCGGACTATCTCGGTGGGATATCGCAGAGGCCAGAG 1560
QY 521 LysIleThrProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 540
DB 1561 AAGATCTGGCCCGAGTACAAGACGATGGAAAGCGGTACACTACGATACGATACGATCGCTCC 1620
QY 541 ValProLeuProThrAlaThrSerMetSerSerValGlyMet 554
DB 1621 GTGCCGTGCTACGGCTACCGATGTCAGTCTCCAGTCTTGGTATG 1662

RESULT 2
AF394242 2245 bp DNA linear PLN 24-JUL-2001
LOCUS Aspergillus oryzae strain TK3 carboxypeptidase S1 (cpi) gene,
DEFINITION


```

complete cds.
ACCESSION AF394242
VERSION AF394242.1 GI:15004615
KEYWORDS
SOURCE
ORGANISM
Aspergillus oryzae
Aspergillus oryzae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 2245)
AUTHORS van den Broek, P.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,
Lausanne CH-1000, Switzerland
FEATURES
Location/Qualifiers
1..2245
/organism="Aspergillus oryzae"
/mol_type="genomic DNA"
/strain="TK3"
/db_xref="taxon:5062"
<1..>2245
/gene="cpi"
join(1..349,409..513,576..692,743..866,941..956,
1008..1124,1187..1237,1296..1449,1522..1580,1670..>2245)
/gene="cpi"
/product="serine carboxypeptidase 1"
join(1..349,409..513,576..692,743..866,941..956,
1008..1124,1187..1237,1296..1449,1522..1580,1670..2245)
/gene="cpi"
/EC number="3.4.16.6"
/codon_start=1
/product="carboxypeptidase S1"
/protein_id="AAK77166.1"
/db_xref="GI:15004616"
/translation="MRGYFELSVLPLVAASWALPGSTAPSVGERQLPKNPTGKILTT
ANNVTRYKEPGAEVCETTRPGVKSXGVDTSPESHFFWFEEAHNPETAPITLWL
NGSPGSDLIIGLFEELGPGCHVNSTDDYINPHSWNEVSNLLFLSQPLGVGFYNPTVD
GSGINPTGVNNSFAGVQGRYPTIDATLIDTTNLAAEAAWEILQFLSGLPSLDSRV
QSKDLSTVESTGYGHYPAFFNHFEYQNERIANGVNGVQLNFNSLGIINGIIDRAIQ
APYTFEFVNTYTGIKAVNETVYVYMKFANQMGCDLITCKOTNPTALADYALCA
EATNWCMDNVEGYPYAFACRGVVDIHPHVDPTPSYKFKLAKSVMDAIGWNINYT
QSNNDVYVAFQOTGDFWENFIEDLEILLALPVRSLIYGDADYICNWFEGQAVSLAA
NYSQAQFRSAGTYELKVNVEYREYGNFSTRVVEAGHEVYQPIASLQLFNR
TIFGMDIAGQGKKIWPXYKNTGATATATHTQSSVPLPTATSMSSVGM"
1..54
sig_peptide
BASE COUNT 548 a 603 c 524 g 570 t
ORIGIN
Alignment Scores:
Pred. No.: 8e-194 Length: 2245
Score: 2618.00 Matches: 548
Percent Similarity: 73.20% Conservative: 1
Best Local Similarity: 73.07% Mismatches: 6
Query Match: 87.88% Indels: 201
DB: 8 Gaps: 9
US-09-712-338-2 (1-555) x AF394242 (1-2245)
QY 1 MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuPro 20
Db 1 ATGAGGGGCTACGAATTTCTCTCAGTGTCTACCCCTTGGTTCAGCCAGTTGGGCCCTTCCA 60
QY 21 GlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGlyVallys 40
Db 61 GGAAGTACACCGCGCTCGGTGAGACAGCTACCCAGAACCCACCAGGGGTCAAG 120
QY 41 ThrLeuThrAlaAsnValThrLeArgTyrLysGluProGlyAlaGluGlyVal 60
Db 121 ACTCTTACAACCGCAACAATGTACCATCCGTTACAGGAACCCGGGCGAGGGCGTC 180
QY 61 CysGluThrProGlyVallysSerTyrSerGlyTyrValAspThrSerProGluSer 80
Db 181 TCGGAGACTACCCCGGGTGTCAAAATCCTACTCTGGATATGTGCACACCTCTCCCGAGTCC 240
81 HisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
241 CATACCTTCTTCTGGTTCTTCGAAGCCAGACATAAACCAGAACTGCACCTATCACATTG 300
101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu 116
301 TGGTTGATGTCGGCCCTCGAAGCGATTCTTTGATCGGTCTCTTCGA-AGTTTGTCTGA 359
117 -----GluLeuGlyPro 120
360 CGCTCGTTACAAGTATGCTCTTTTAGTCTCGTTAACTTGGTTAGAGTTGGGCCCT 419
121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 140
420 TGCATGTCAATTCGACTTTTGTGATGACTACATCAACCTCACTCGTGAACGAGGTCTCC 479
141 AsnLeuPheLeuSerGlnProLeuGly 150
480 AATTACTATTCTGTCCAGCCATT-GGAGTCGGTATGATGCGGCACCCCTCCCTA 538
151 -----ValGlyPheSerTyrSerAspThrVal 159
539 CTGGGCTCTTCCCATATTGACGAGGTGCTTCCCGTAGGCTTTTCATATAATGATACGGTT 598
160 AspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGln 179
599 GATGGGTCCATTAAACCTGTAACTGGGTGCTGAAAAATTCGAGCTTTGCAGGAGTTTCA 658
180 GlyArgTyrProThrIleAspAlaThrLeuIle 190
659 GSCCGGTATCCCAACCAATTGATGCCACTCGAT-CGGTAAACGGTTTCTCCGCTCCCATG 717
191 -----AspThrThrAsnLeuAlaAlaGluAlaAlaTrpGlu 202
718 CTCTATTATTACTGATACGATAGCAGATACTACCAATCTTCCGCCAGAGCCGCTTGGAG 777
203 IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp 222
778 ATCTCTGCAAGGATTCCTTAGTGGACTTACCTAGCTTGGACTTAGGTCTAGGTCTAAGGAC 837
223 PheSerLeuThrGluSerTyrGlyGly 232
838 TTCAGTCTATGACGAGAGCTATGGAGG-GTAGTGGTTCCTCTCGAAATCTCTTAA 896
233 -----HisTyrGlyProAla 237
897 CGGCCGTGATGCTTGGCTGATAGGTTTCTTTCGTTTCGAAAGGCACATATGCTCTGCA 956
238 -----PhePheAsn 240
957 GTGAGTTCTACTGGAGCATGTGTCAAAGAGGCAAGACTAACAGAACGCTAGTTCTTCAAT 1016
241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
1017 CATTTTACGACAGAAATGCCAAGATTCGACGAGTGTGTAAATGTTGTTTCAAGTTAAT 1076
261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAla 277
1077 TTCAACTCTCTGGGAATTTATAACGCGCATCATCGACGAGGCGATCCAGT-ATACCGTCG 1135
278 -----ProTyr 279
1136 AAACAGCCCATAGAACATCTCTCCGAAACAAAGCTGCTAACGATATGCCATAGGCCCTTAC 1195
280 TyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAla 293
1196 TACCTGAAATTCGTGTGTGAACAATACCTACGCTATACGCTATACAGGCTGTAAAGTTCC 1255
294 -----ValAsnGluThrValTyrAs 300
1256 GCATGAGATTATATCTCTGTGATCTCATCCATGAAAGGTCAACGAGACCGTCTACAA 1315

```

QY	300	nTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuLeuSerThrCysLy	320	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Db	1316	CTACATGAGTTTCACCAACCAATGTCACAAATGGTTGCAGGATTTGATTTCCACTCGCA	1375	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
QY	320	sGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCy	340	Edens, L., dijk Van, A. A., Krubasik, P., Albermann, K., Stock, A.,
Db	1376	ACAGACAAACCGCAGCGCATTTAGCTAGCTACGCCCTCTGCGCCGAGCCACCAACATGTG	1435	Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,
QY	340	sArgAspAsnVal	344	Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,
Db	1436	CAGGACAAATGTTGGTAGCTTCTTTTCCCTCTTGGAAACAAAGGTATATGATTCAAAT	1495	Kemmer, W., Tan, P., Stiebner, J. and Albang, R.
QY	345	-----GluGlyProTyrTyrAlaPheAlaGlyArgGlyVa	356	Novel genes encoding novel proteolytic enzymes
Db	1496	AGTGGGACAATAACACGAAAAATAGAGGGGCCATACACCTTTTGTGCTGGTGTGT	1555	Patent: WO 02068623-A 109 06-SEP-2002;
QY	356	lTyrAspIleArgHisProTyrAsp	364	DSM N.V. (NL)
Db	1556	GTATGATATTCGGCATCCATATGA-TGTAAGTAGAGCTTTTACTTCTATTTACACCGC	1614	Location/Qualifiers
QY	365	-----AspPr	366	1. 1656
Db	1615	ATGCAAATGCATGTTTCATTGTTGATCGAATAAAATCCCCCTAATATGTTAGGACCC	1674	/organism="Aspergillus niger"
QY	366	oThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGl	386	/mol_type="genomic DNA"
Db	1675	GACTCCGCCAAGTATTACACAAATTTCTGCCAAGGACTCTGTCTATGAGCGCTATCGG	1734	/db_xref="taxon:5061"
QY	386	yValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnThrGl	406	BASE COUNT 411 a 426 c 409 g 410 t
Db	1735	CGTCAACATCACTACACCCAGTCCCAATATGACGCTCTACTACGCTTTCCAGCAACAGG	1794	ORIGIN
QY	406	yAspPheValTyrProAsnPheIleGluAspLeuGluLysLeuAlaLeuProValAr	426	Alignment Scores:
Db	1795	CGACTTTGCTGGCCCACTTCATCGAGACCTCGAGGATCCCTTGTCTCCCGTGGC	1854	Pred. No.: 3.35e-175 Length: 1656
QY	426	gValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlnAlaVa	446	Score: 2376.00 Matches: 428
Db	1855	TGTCCTCCCTCATCTATGGCGACCGCGATTACATCTGCAACTGGTTCGGCGGTGAGCGCT	1914	Percent Similarity: 87.96% Conservative: 54
QY	446	lSerLeuAlaAlaAsnTyrSerGlnAlaGlnPheArgSerAlaGlyTyrThrProle	466	Best Local Similarity: 78.10% Mismatches: 66
Db	1915	TTCCCTCGCTGGAACTACTCCAGCGCCCGAGTTCCGAGCGGATGATTCCTCTCACTCG	1974	Query Match: 79.76% Indels: 0
QY	466	uLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrAr	486	DB: 6
Db	1975	GAAAGTCAACGGCTCGAGTATGGGAACTCGGAGTATGTAATTTCTCTCACTCG	2034	US-09-712-338-2 (1-555) x AX534871 (1-1656)
QY	486	gValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPh	506	QY 1 MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuPro 20
Db	2035	CGTCTATGAGGAGGCGCATGAAGTCCCTACTTACCAGCCCATCGCTCCCTCGCAATTGT	2094	Db 1 ATGCGTGGCTCGGTTGGTCTTGTGTGCCCCCTGGTGCATTTAGTTGTCTATGCC 60
QY	506	eAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLysLysIleTrpProSerTy	526	QY 21 GlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGlyValLys 40
Db	2095	TAACCGGATCTCTCGGTTGGATATCGAGAGGCGCAGAGAAATCTGCCCCAGCTA	2154	Db 61 GAGAAATGATGTCATCTACGATAAGAGGAGGTACCAAAAGCGCTCCACTCGCGTCAA 120
QY	526	xLysThrAsnGlyThrAlaThrHisThrGlnSerSerValProLeuProThrAl	546	QY 41 ThrLeuThrThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal 60
Db	2155	CAAGAGCATGGACCGCTACAGTACGATACACATACAGTGTGCTGCGCTGCTACGGC	2214	Db 121 TCGATAAAACCCCAACATGTCATCAGGTATAAGAACACAGAACCGAGGAANT 180
QY	546	aThrSerMetSerSerValGlyMetAla 555		QY 61 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer 80
Db	2215	TACCAGCATGTCCAGTGTGGTATGGCT 2242		Db 181 TGTGAGACACACCTGGGGTCAAATCATCTCCGATATGTCGATCTTTCGCCACAGATCG 240
RESULT 3				QY 81 HisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
AX534871				Db 241 CATACTTTCTTTGTTTTCGAGTCACGCCGTGACCCGAAATCATCCAGTACTCTG 300
LOCUS	AX534871	1656 bp	DNA	QY 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 120
DEFINITION	Sequence 109 from Patent WO02068623.			Db 301 TGGCTGAATGTTGGGCCCTGGAAAGCGATTCCTTGATTGGCTTTTGAAGAGTTGGTCCG 360
ACCESSION	AX534871			QY 121 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSer 140
VERSION	AX534871.1	GI:25261462		Db 361 TGTCAATCACACACAGAGTACCAATCATCATCATCTCTCTGGAACAGAGTCCACC 420
KEYWORDS	Aspergillus niger			QY 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
SOURCE	Aspergillus niger			Db 421 AATCTTCTTTCTTGTCTCAGCCCTCGGTGGGGTCTCTTCTACAGTGAACCCGAGGCC 480
ORGANISM	Aspergillus niger			QY 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
				Db 481 GGGTCTCTTGAAATCCATTTACTGGAGCCGTCGAGAGCCCTCTCTTCTGCTGAGTTTCAAGGT 540
				QY 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla 200
				Db 541 CGATACCCAGTATTATGATCCATCATCATCGACACGCGATATCGCTGCACCGCAACC 600
				QY 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
				Db 601 TGGGAGGTGCTTCAGGGCTTCCTCAGTGGCCCTGTCAGCTAGTATCCCGAGTCAAGTCC 660
				QY 221 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsn 240

Db	661	AAGGAGTTTCAACCTGTGCACAGAGATTACGGAGGACACTATGGACCAGCGTTCTTCAAT	720
Qy	241	HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn	260
Db	721	CATTCTACGACGAAAATTCGAAGATCGTAGCGGGGAAGTCAATGGCGTCCACTGAAT	780
Qy	261	PheAsnSerLeuGlyIleAlaAsnGlyIleAlaAspGluAlaIleGlnAlaProTyrTyr	280
Db	781	TTTAACTCCCTCGGAGTTATCAACGGCATCATTTGATCGCGGATTTCAGCGAGACTAC	840
Qy	281	ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn	300
Db	841	GCAGACTTTGCGGTATAATACATATGGAATCAAAGCTGTCAATGACACAGTGTACAC	900
Qy	301	TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys	320
Db	901	TATATGAAGTTTCGCCAACACCATGCGAAATGGATGCCAGGATCATGGTTCTCGTGTAAA	960
Qy	321	GlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys	340
Db	961	TTGACCAATAGACCTCGCTTCTGATTATGCTATATGTACAGAAGACGCCAATATGTGC	1020
Qy	341	ArgAspIleValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg	360
Db	1021	AGGACAAATGTGGAAGGCGCTTACTACCAAGTTTGGCGCGCTGGCTGTATGATATTCG	1080
Qy	361	HisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSer	380
Db	1081	CACCCCTACAAATGACCCGACCCCGCGCTCTTGTGTACTACCTCAAGAAAGACTCA	1140
Qy	381	ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnValTyrTyr	400
Db	1141	GTCAATGATGCTATCGCGGTGGACATTAACTACACCGAGTCCAGCGCGAATATAT	1200
Qy	401	AlaPheGlnGlnThrGlyAspPheValTyrProAsnPheIleGluAspLeuGluGluIle	420
Db	1201	GCATTCAGCAGACCGCGCGACTTTGTATGCGCGAAITTCATTGAGGACCTCGAAGAGATC	1260
Qy	421	LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp	440
Db	1261	CTCCAACTCCCGGTACCGGTGTCGTTGATCTACGGCGATGCGGACTATATCTGTAACTGG	1320
Qy	441	PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaGlnPheArgSer	460
Db	1321	TTGCGCGGTGAGGCAATCTCACTCGCAGTTAACTACCCCATCGAGTCTAGTCCGTGCA	1380
Qy	461	AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly	480
Db	1381	CGCGGATACACCCATGACAGTAGATGGGTGGAATACGGTGAGACTCGCGAGTATGGC	1440
Qy	481	AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle	500
Db	1441	AACTTTTCGTTCCACCGGTATATCAGGCTGGGACAGAGTTCATATCATCATCAACCGATC	1500
Qy	501	AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys	520
Db	1501	GCAGCGTTCAGCTGTTCAACCGTACTTTATTTGGATGGGATATTGACGCGGTACAAC	1560
Qy	521	LysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer	540
Db	1561	CAGATTTCGCCCCGAATATAGCACCAACGGGACATCGAGGCTACACACGGAGTCTTC	1620
Qy	541	ValProLeuProThrAlaThrSer	548
Db	1621	GTGCCACTGTCCACGGCGTCGAGT	1644
RESULT 4			
AX534814			
LOCUS			
AX534814			
SEQUENCE 52 from Patent WO02068623.			
AX534814			
VERSION			
AX534814.1 GI:25261343			
LINEAR			
PAT 22-NOV-2002			

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT
ORIGIN
Alignment Scores:
 Pred. No.: 2.78e-147 Length: 3150
 Score: 6231.00 Matches: 421
 Percent Similarity: 64.97% Conservative: 52
 Best Local Similarity: 57.83% Mismatches: 75
 Query Match: 67.84% Indels: 185
 DB: 6 Gaps: 8
 US-09-712-338-2 (1-555) x AX534814 (1-3150)
 QY 1 MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTyrAlaLeuPro 20
 Db 575 ATGGCTGGCTCTCGATTGGTGTCTTCTTGCCTCGCTGCTGCTGCTATGTTGGCTATGCC 634
 QY 21 GlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGlyValLys 40
 Db 635 GAGAAATGAATGGTCACTACGATAAGAAGCAGGTACCAAAGCGTCCACTGGCGTCAAA 694
 QY 41 ThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal 60
 Db 695 TCGATAAAACCCCAACATATGTCATATCAGGTATTAAGAACGAGAACCGAAGGAATT 754
 QY 61 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer 80
 Db 755 TGTGAGACACACCTGGGTCAAATCATATCTCGGATATGTCGATCTTTCGCCAGAGTCG 814
 QY 81 HisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
 Db 815 CATACITTCITTTGGTITTCGAGTCACGCCGTCAGCCCGAAATGATCCAGTGACTCTG 874
 QY 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu----- 116
 Db 875 TGSCCTGAATGGTGGCCCTGGAGCGATTCTTGATTGGCTTTTGA-AGGTTGGCCAAA 933
 QY 117 -----GluLeu 118
 Db 934 TATCCTGACGAAAAGATAAAATTCAGTTGCATGTTCTGACGCTTCACACAGAGTTG 993
 QY 119 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu 138
 Db 994 GGTCCGTGTACATCACACACGAGTACGAATCAATCATCAATCAGTACTCTCTGGAACGAG 1053
 QY 139 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyVal----- 151
 Db 1054 GTACCAATCTCTCTTCTTGCTCTAGCCCTCGGTGGGTATGGAATATTGCTGCTT 1113
 QY 152 -----GlyPheSerTyrSerAspTh 158
 Db 1114 CATACATCTGAGTACATTGCTTATCGGTCTTATCTCGGAAGGGTTCTCTTACAGTGAAC 1173
 QY 158 rValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVa 178

Db	1174	CGAGCGGGTCTCTGAATCCATTACTTGGAGCGCTCGAGAACGCTCTCTTGTGCGAGT	1233
Qy	178	lGlnGlyArgTyrProThrIleAspAla	187
Db	1234	TCAGGGTCGATACCCAGTATTGATGACCACTATCATCGTGAAGTTGCCGGTTGACTCT	1293
Qy	188	-----ThrLeuIleAspThrThrAsnLeuAlaAlaGluAl	199
Db	1294	CACCTAGCATCTCTCAATCTCTACTTTACA-GACACGACCGATATCGTCGACGCGC	1352
Qy	199	aAlaTyrGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGl	219
Db	1353	AACCTGGAGGTGCTTACGGGCTTCTCAGTGGCTGTCGAGCTAGATTCCGAAGTCAA	1412
Qy	219	nSerIysAspPheSerLeuThrThrGluSerTyrGly-GlyHisTyrGly	235
Db	1413	GTCCAAAGGAGTTCAACCTGTGGACAGAGATTACGGAGGGTGAGTGCAACTTTCATACA	1472
Qy	235	-----	235
Db	1473	GACCGACGTAAAGCTGACTTGATCAAGACACTATGACACCGGTAGTGTCTTTCTCG	1532
Qy	236	-----ProAlaPheAsnHisPheTyrGluGlnAsnGluA	248
Db	1533	ITGCACACATATTGATCTAATGACCGAAGTCTTCAATCATTTCTACGAGCAAAATTCCA	1592
Qy	248	rgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleA	268
Db	1593	AGATCGTAGCGGGGAAGTCANATGGCTCCACTGATTTTAACCTCCCTCGGGNTATCA	1652
Qy	268	snGlyIleIleAspGluAlaIleGlnAlaPro	278
Db	1653	ACGGCATCTTATGTCGCGGATTTCAGGTACT-TAGAAATGACGCTCGCGACAGGCTGG	1711
Qy	279	-----TyrTyrProGluPheAlaValA	286
Db	1712	GCCTAGAGGACATCGCTAAAGTAAATTAATAGCAGACTACTACGAGACTTTGCCGTTA	1771
Qy	286	snAsnThrTyrGlyIleLysAla	293
Db	1772	ATAATACATATGAATCAAGCTGAAGTTAAATACAGTACATCGTGGATTAAGATC	1831
Qy	294	-----ValAsnGluThrValTyrAsnTyrMetIysPheAlaAsn	306
Db	1832	AACCGTGCTCATGCTTGCTAGGTCAATGACACAGTGTACAACTATATGAAGTTCCGCAAC	1891
Qy	307	GlnMetProAsnGlyCysGlnAspLeuIleSerThrCysIysGlnThrAsnArgThrAla	326
Db	1892	ACGATCCCAATGGATGCCAGGATCAGGTGCTTCGTGTAATTGACCAATAGGACCTCG	1951
Qy	327	LeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnVal	344
Db	1952	CTTCTGATATGCTATATGTACAGAGGACCAATATGTGCGGACCAATGT-CGGTGA	2010
Qy	344	-----	344
Db	2011	GTGGTTCTACTGTTCTCTCGAGGGTGCAATGATGAAGGACTTTGCTAAGCTGTATGT	2070
Qy	345	---GluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr	363
Db	2071	ACAGAGGGCTTACTACCAAGTTTGGCGCGGTGGCGTGTATGATATTCGGCACCCCTAC	2130
Qy	364	Asp	364
Db	2131	AA-TGTAAGTGCGAAGGATAAGGATTGCTCTTCGGAACAGGACACTGCTCATATGTCA	2189
Qy	365	---AspProThrProProSerTyrTyrAsnIysPheLeuAlaIysAspSerValMet	382
Db	2190	ACGTAGGACCCCGCCGCTCTACTTGTGTTGATACCTCAAGAAAGACTAGTCATG	2249
Qy	383	AspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyrAlaPhe	402
Db	2250	GATGCTATCGCGTGGACATTACTACACCGAGTCCAGCGCGGCGAAGTATATTATGCATTC	2309
Qy	403	GlnGlnThrGlyAspPheValTyrProAsnPheIleGluAspLeuGluIleLeuAla	422
Db	2310	CAGCAGACCGCGGACTTTTGTATGGCCGAATTTCAATTGAGGACCTCGAAGAGATCTCCAA	2369
Qy	423	LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGly	442
Db	2370	CTCCCCGTACGCGTGTGTTGATCTAGCGGATCCGACTATATCTGTAACTGGTTCGGC	2429
Qy	443	GlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGly	462
Db	2430	GSTCAGGCCATCTCACTCGAGTTAACTACCCCATCGAGCTCAGTTCCGTGCGACGGGA	2489
Qy	463	TyrThrProLeuIysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPhe	482
Db	2490	TACACACCATGACAGTAGATGGGTGGAATACCGTGAGACTCGGAGATGATGGCAACTTT	2549
Qy	483	SerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSer	502
Db	2550	TCGTTTACCCCGGTATATACGGCTGGGACGAGGTTCCATCTATCAACCGATCGACGG	2609
Qy	503	LeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLysIle	522
Db	2610	TTGACGCTGTTCAACCGTACTTTATTGATGGGATATTGCGCGGTACACTCAGATT	2669
Qy	523	TrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValPro	542
Db	2670	TGGCCGAATATAGCACCAACGGGACATCGCAGGCTCACACACGAGTGTGTTGTTGCCA	2729
Qy	543	LeuProThrAlaThrSer	548
Db	2730	CTGTCCACGGCTCGAGT	2747
RESULT 5			
LOCUS	AX534872	1872 bp	DNA linear PAT 22-NOV-2002
DEFINITION	Sequence 110 from Patent WO02068623.		
ACCESSION	AX534872		
VERSION	AX534872.1	GI:25261464	
KEYWORDS	Aspergillus niger		
SOURCE	Aspergillus niger		
ORGANISM	Aspergillus niger		
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.		
AUTHORS	Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A., Kimpel, B., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W., Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiebler, J. and Albang, R.		
TITLE	Novel genes encoding novel proteolytic enzymes		
JOURNAL	Patent: WO 02068623-A 110 06-SEP-2002;		
FEATURES	DSM N.V. (NL)		
source	Location/Qualifiers		
	1..1872		
	/organism="Aspergillus niger"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:5061"		
BASE COUNT	402 a 434 c 570 g 466 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	8 83e-67	Length:	1872
Score:	980.50	Matches:	216
Percent Similarity:	56.23%	Conservative:	91
Best Local Similarity:	39.56%	Mismatches:	178
Query Match:	32.91%	Indels:	61
DB:	6	Gaps:	19
US-09-712-338-2 (1-555) x AX534872 (1-1872)			
Qy	44	ThrAlaAsnAsnValThrIleArgTyrLysGlu	-----ProGlyAlaGlu----- 58
Db	94	ACGCGTGGAGGATCTCAACCGTATTTCATTCGAGATATTCCTGTCGAGGATCTCTAT	153

QY 59 -----GlyValCysGluThrThrProGlyVallySerTyrSerGlyTyrVal 74
Db 154 AAGCAACCCCTCGGCATCTGCACCAACCCCTCCACCCCGAGTACTCCGGCTACATC 213
QY 75 AspThrSerProGluSer----- 80
Db 214 CACCTCCCCCAACACACCCCTTACCAATCTCTCCATTCAGGAATCAGCATCTCGCAACA 273
QY 81 -----HisThrPheThrPhePheGluAlaArg-----HisAsnProGluThr 95
Db 274 TACCTATCAATACCTTTCTGTACTTTCTCCCGCCATACCAACAAATGATACA 333
QY 96 AlaProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPhe 115
Db 334 TCCCACTCACTCACTCGATGATGAACCGCGCGGCCCGCGGATCCTCCATGATGGGCTATT 393
QY 116 GluGluLeuGlyProCysHisValAsnSerThrPheAspAsp-----TyrIleAsnProHis 134
Db 394 CAGAGAACCGGGCCATCTACTGTGAATACGAGCTCGAATCCACGGCTATAATCCCTGG 453
QY 135 SerTyrAsnGluValSerAsnLeuPheLeuSerGlnProLeuGlyValGlyPheSer 154
Db 454 TCGTGGATGATGATCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 513
QY 155 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer 174
Db 514 TATGATGTGTGAGGAATGGACGCTTAGAT-----TTGAATGAGACG 555
QY 175 PheAlaGlyValGlnGlyArgTyrProThrIleAsp-----AlaThrLeuIleAspThr 192
Db 556 TTT-----TTGTGGGAGCGTGTCCGAGTCAGGATGTGATGCGGCGGTGAATGGGACG 609
QY 193 ThrAsnLeuAlaAlaGluAlaAlaTyrGluLeuGlnGlyPheLeuSerGly----- 210
Db 610 GTTAAT---GGGGAAGGGCGCTTTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 666
QY 211 -----LeuProSerLeuAspSerArgValGlnSerLysAsp---PheSerLeuTyrThr 227
Db 667 GAATATGTTCTCTGTGACGGGAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 726
QY 228 GluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnGlu 247
Db 727 GAGTCATATGGGGGAGCGGTATGACCGGCATACACCGCGCTCTTCAGGAGATGAATGAG 786
QY 248 ArgIleAlaAsnGlySerValAsn---GlyValGlnLeuAsnPheAsnSerLeuGlyIle 266
Db 787 AGGATTGAGTGGGAGGTAAACACCGGGAAGAGATCCATTTGATAGCTGGCGCAT 846
QY 267 IleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsn 286
Db 847 ATCAATGGGTGTGGATTTACTCGTGCAGTCCCTTCCTTCCTCGAGCAGCGGTATAAC 906
QY 287 AsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsn 306
Db 907 AATACGTATGGGATCGAGGGAATCAATCGCAGCTCTAGACCGGCTATGATAGTTGG 966
QY 307 GlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAla 326
Db 967 AGCAACCTCGCGGTGCGAGGATATGATCATGAGTGTGGGATGCTGGCGAG----- 1020
QY 327 LeuAlaAsp---TyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGlu 345
Db 1021 CTCGGAGATCCCTCATCTCTCGAGGAGGCGTCCGACTACTGTTCGGCGGAGATCAAG 1080
QY 346 GlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAsp 365
Db 1081 AGCTGTATACGAATACCTCCGGCGGAGATACATGACATGACGATTCACGCGCGAT 1140
QY 366 ProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIle 385
Db 1141 GCAGCTCTCGTCTTACTTCGTGGGTCTTGAATCGCCCATGGGTGCAAAAGGCACTT 1200

QY 386 GlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnGlnThr 405
Db 1201 GGGTCCCGGTGAATATACCATCTCGTCAAGGCGAGTGGGAAACAGTTCCTCGCTCGAG 1260
QY 406 GlyAspPheValTyrPro-----AsnPheIleGluAspLeuGluGluIle 420
Db 1261 GGCATTTAT-----CCGCGAATGATCCCGCGGAATGATCGGGGATATTGGATCTTG 1314
QY 421 LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyr 440
Db 1315 CTTGACTCCGGTGTCAAGTGGCTATGATATATGGGACCGGAGTATGCTTGTTCCTGG 1374
QY 441 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 460
Db 1375 CGCGCGCGGGAAGATGTCAAGCTCTCTGTGGAGTACGAGGATGCGGAGAAGTCCGTGCT 1434
QY 461 AlaGlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGlu 478
Db 1435 GCTGGGTATCCGAGTGCAGACGAGTCA---TCCTACGTGGGGTCTAGTAAGCGAG 1491
QY 479 TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln 498
Db 1492 TATGGGAATCTCGTTCAAGCGTCTTTTTCAGCGGCGCATGAGTGGCCATTTATTCAG 1551
QY 499 ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGly 518
Db 1552 CCCGAAACGGCGTATGAGATTTTAAATCGCGCTCAGTTTAAATGGGATATTTCGACGGA 1611
QY 519 GlnLysLysIleTyr-----ProSerTyrLysThrAsnGlyThrAlaThrAlaHis 536
Db 1612 GGCATTTCTCGAGCAGATCAGAGCTATGGACGGGAGCGCTCGTCAAGTGGCAT 1671
QY 537 ThrGlnSerSerValPro 542
Db 1672 ATCAAAAACGAGTGGCG 1689
RESULT 6
AX534815
LOCUS AX534815 3221 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 53 from Patent WO02068623.
ACCESSION AX534815
VERSION AX534815.1 GI:25261346
KEYWORDS Aspergillus niger
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
REFERENCE 1 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
AUTHORS Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A., Kimpe, B., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W., Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiebler, J. and Albarg, R.
TITLE Novel genes encoding novel proteolytic enzymes
JOURNAL Patent: WO 02068623-A 53 06-SEP-2002;
FEATURES
source location/Qualifiers
1. 3221
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT 739 a 773 c 869 g 840 t
ORIGIN
Alignment Scores:
Pred. No.: 3-29e-66 Length: 3221
Score: 977.00 Matches: 213
Percent Similarity: 55.68% Conservative: 90
Best Local Similarity: 39.08% Mismatches: 172
Query Match: 32.80% Indels: 70
DB: 6 Gaps: 17
US-09-712-338-2 (1-555) x AX534815 (1-3221)

QY 59 GlyValCysGluThrThrProGlyValIysSerTyrSerGlyTyrValAspThrSerPro 78
DB 918 GGCATCTGCACACCCCTCCACCCCGAGTACTCCGGCTACATCCATCCCTCCCGCA 977
QY 79 GluSer-----His 81
DB 978 CACACCCCTACCAATCTCTCCATCCAGGAATCAGATCTCCGACCAACATCCATCAAT 1037
QY 82 ThrPhePheThrPheGluAlaArg-----HisAsnProGluThrAlaProIleThr 99
DB 1038 ACCCTTTTCTGGTACTTCTTCCCGCCATCACCACACATGATACATCCCTCCACTACC 1097
QY 100 LeuThrLeuAsnGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGly 119
DB 1098 ATCTGGATGAACCGCGCGCGCGGATCTCCATGATGGGCTATTTCAGAGAACCGG 1157
QY 120 ProCysHisValAsnSerThrPheAspAsp---TyrIleAsnProHisSerTrpAsnGlu 138
DB 1158 CCATGTACTGTGAATACGAGCTCGAATCCACGGCTATATCCCTGGTGGTGGATGAG 1217
QY 139 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 158
DB 1218 TACGTTCGATATGTTATATTGAGCAGCGCGGTGCAGACGGGATTTAGTTATGATGTTG 1277
QY 159 ValAspGlySerIleAsnProValThr----- 167
DB 1278 AGGAATGGACGCTTAGATTGGTTAGTGGGAGATAGATGTTAGTATTAGTGTGAG 1337
QY 168 -----GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIle 185
DB 1338 AGGGATGGAGTAGGACAGATGAGCGTTT-----TTGGTGGGACGTTGCCGAGTCAG 1391
QY 186 Asp-----AlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIle 203
DB 1392 GATGTCATGGACCGGTGATGGACGCTTAAT---GGGGAGGCGCTTGGGTGCG 1448
QY 204 LeuGlnGlyPheLeuSerGly-----LeuProSerLeuAspSerArgValGln 219
DB 1449 TTGCAGGTTTGGTGGGTGAATCTCTGAATATGTTTCTCTGTTCGACGGGAATGTTGT 1508
QY 220 SerLysAsp---PheSerLeuThrThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 238
DB 1509 GGTGATGACAGGCTGATGATGAGCGAGTCAATATGGGACGCGTATGACCGCGGATAC 1568
QY 239 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsn---GlyVal 257
DB 1569 ACGGCGCTCTTCAGGAGATGATGAGAGGATTCAGAGTGGGAGGTAAGCACCAGGAAG 1628
QY 258 GlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAla 277
DB 1629 AAGATCCATTTGGATACGCTGGGCAATTATCAATGGGTGTGTGGATTACTCGTGCAAGTC 1688
QY 278 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThr 297
DB 1689 CCTTCGTCTCCTGAGCGGCTATACAAATACGATGCGGATCGAGGAATCAATCGCACG 1748
QY 298 ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 317
DB 1749 CTCTACGACCGGCTATGATAGTTGGAGCAAGCCTGGCGGTTCAGGGATATGATCATC 1808
QY 318 ThrCysLysGlnThr-----AsnArgThrAla 326
DB 1809 GAGTGTCCGATGTCGCGAGCTCGGAGATCCCTCAATGATGGCGCAATGACCGGTA 1868
QY 327 LeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGly 346
DB 1869 -----AATACATCTCGAGAGCGCTGCGACTACTGTTCCGCGGAGATCAAGAGC 1919
QY 347 ProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro 366
DB 1920 CTGTATACGAATACCTCCGGCGGAGGATCTACGACATAGCGCATTTTCAGCCCGGATGCA 1979
QY 367 ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGly 386

DB 1380 GCTCTCGTCCCTTACTTCTGTCGGGTCTTGAATGCCCATCGGTGCAAAAGCACTTGGG 2039
QY 387 ValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnGlnThrGly 406
DB 2040 GTCCCGGTGAATATACCATGCTGCTCAGAGGAGTGGGGAACAGTTTCGCTCGACGGGC 2099
QY 407 AspPheValTrpPro-----AsnPheIleGluAspLeuGluLeuLeu 421
DB 2100 GATTAT-----CCGCGAATATGATCCCGCGAATGATCGGGGATATTGGATACTGCTT 2153
QY 422 AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 441
DB 2154 GACTCCGGTGCAGGTGGCTATGATATATGGGACCGGACTATGCTTGTCCGTGGCGC 2213
QY 442 GlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 461
DB 2214 GCGGGGAAGATGTGAGCTGCTGAGTACGAGGATGCGGAGATTCGCTGCTGCT 2273
QY 462 GlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGluTyr 479
DB 2274 GGGTATGCCAAGTGCAGACGAAAGTCA---TCCTACGTTGGGGGCTCTAGTAAGGCGAT 2330
QY 480 GlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnPro 499
DB 2331 GGGAACTTCTGCTTCCGCTGCTTTCAGCGGCGCATGAGTGCCTATTTATCAGCCC 2390
QY 500 IleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGln 519
DB 2391 GAAACGGCTATGAGATTTTAAATCGGCTCAGTTTAAATGGGATATTCGACGGGAGC 2450
QY 520 LysLysIleTrp-----ProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThr 537
DB 2451 ATTCTCTGACGACGAATCAGAGCTATGGACGGGAGCCGCTGTCACACGTTGATATC 2510
QY 538 GlnSerSerValPro 542
DB 2511 AAAACGAAGTCCCG 2525
RESULT 7
LOCUS AX534866 1665 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 104 from Patent WO02068623.
ACCESSION AX534866
VERSION AX534866.1 GI:25261452
KEYWORDS Aspergillus niger
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
REFERENCE 1 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
AUTHORS Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
TITLE Novel genes encoding novel proteolytic enzymes
JOURNAL Patent: WO 02068623-A 104 06-SEP-2002;
DSM N.V. (NL)
FEATURES Location/Qualifiers
source 1. 1665
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT 374 a 455 c 424 g 412 t
ORIGIN
Alignment Scores:
Pred. No.: 4,75e-64 Length: 1665
Score: 944.50 Matches: 214
Percent Similarity: 51.52% Conservative: 75
Best Local Similarity: 38.15% Mismatches: 176
Query Match: 31.71% Indels: 96
DB: 6 Gaps: 14

US-09-712-338-2 (1-555) x AX534866 (1-1665)

QY	14	AlaAlaSerTrpAlaLeuProGlySerThrProAlaSerValGlyArgArg-GlnLeuPr	33
Dd	12	TCTGTCTTGGGGTCTTTGGGTCTAGCGAC-----CCTCAATTTC	56
QY	33	pLysAsnProThrGlyIleHisThrLeuThrThraAla-----AsnAsnValThrIleAr	51
Dd	57	TCCCGAGCGGAAGGCATCATCTGCTCAAGTCCAAGTCCATGAGAAATGTAATTTTC	116
QY	51	GTYrlysGluProGlyAlaGluGlyValCysGluThrThrProGlyVallysserTyrSe	71
Dd	117	TTTCAAAGAGCT-----GGAATTTGCCAATACCGCGGTGTCGATCTTATTC	167
QY	71	rGlyTyrValAspThrSerProGluSerHisThrPhePheTrpPhePheGluAlaarghi	91
Dd	168	GGGCTATGTACACCTTCCCCCGGCTCAACGAGCTTCTTTTGTTTTCGAAGCCGCA	227
QY	91	sAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLe	111
Dd	228	AGATCCCAGCAATCGGCTCTGGGCATCTGGCTCAATGGCGGTGGGTGGCTCGCT	287
QY	111	uileGlyLeuPheGluGluLeuGlyProCysHisValAsnSer---ThrPheAspTyr	130
Dd	288	CATGGGCTCTTGAAGAATTAGTCTCTTGTTCATTTGCATGACACTCAAGACCACAGT	347
QY	130	rIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGl	150
Dd	348	CCTCAATCCTTGGAGTTGGAACAATGAAGTCAATCTTCTATCTTGACAGCCACTCA	407
QY	150	yValGlyPheSerTyrSerAspThrValaspGlyserIleAsnProValThrGlyValVa	170
Dd	408	AGTCGGCTTCTCAGACGATGCCAACAAATGGCACTTTGACA-----	450
QY	170	LgIuAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIl	190
Dd	451	-----GC	452
QY	190	eAspThrThrAsnLeuAlaGluAlaLaTrpGluIleLeuGlnGlyPheLeuSerGl	210
Dd	453	TAATGGAGCGCATCGCGCTCACGCTCTATGGCATTTTCGCGCAACCTGTTTTTCA	512
QY	210	yLeuProSerLeu-----AspSerArgValGlnSerLysSerAspPheSerLeuTrpTh	227
Dd	513	GTTCACACACTACAGCCAAACGATCATCTGTC-----AGTCTCTGGGC	557
QY	227	rGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGlnAsnGl	247
Dd	558	TGAAGATTACGAGGCGCATATGCTCAGGCATCTTTCGTTCTTCCAACAGCAGATGA	617
QY	247	uAtgIleAlaAsnGlySerVal--AsnGlyValGln--LeuAsnPheAsnSerLeuGl	265
Dd	618	CAAAATCGAGAGGCGGCTCAGAGACGGTGCACAGTATTTGCATCTCAGACGCTTG	677
QY	265	yIlelleAsnGlyIlelleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaVa	285
Dd	678	CATTGTGAACGGCTTGATGATATGTTGATTCAGAGAGGCTTACATTCTTGCCA--	735
QY	285	lAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPheAl	305
Dd	736	-----TACAATAACGTAAGGCTCGC	755
QY	305	aAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgTh	325
Dd	756	CCCT-----TCTTCAATCAACTCGCG	776
QY	325	rAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp-----	342
Dd	777	AGGCTTTTCGGATCAGGCGCTCGCTCGAGAGCGGCTTTTGAAGAAGACGGATTCGGCTT	836
QY	343	-----AsnVal-----GI	345
			:

Db	837	GCCTCACTCAGGGAAGAATATCTCTGAAATTTCCGAGGCGCTTGCATAGATGGGGAGA	89
QY	345	uGlyPro-----TyrTyralaPheAlaGlyArgGlyValtyrAspIleArgHisProTy	363
Dd	897	TGGCCCATCACTTACTACACACCTTCAATCGGGGTGATACGATATCCCATCTCTAA	956
QY	363	rAspAspProThrProSerSerTyrTyralaPheLeuAlaAspSerValMetAs	383
Dd	957	GAAAGCAATTCCTCCAGCACATCTCGGATATTTGACGAGGAGTCCGCTCTTGC	101
QY	383	pAlaileGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValtyrTyralaPheGl	403
Dd	1017	CGCTTTGGGGTACCATGTAATTTCAATCGTCTTCGAGTGGCGGTGCTACACAGTTAT	107
QY	403	nGlnThrGlyAspPheValTrpProAsnPhellegluAspLeuGluGluIleleuAlaLe	423
Dd	1077	AAAAACCTTTGATATCGTCCAGCGCGCTTCTCGGATGCAATTTGGCTACTCTCTGACAG	113
QY	423	uProValArgValSerLeuIleTyrGlyaspAlaAspTyrIleCysAsnTrpPheGlygl	443
Dd	1137	TGGTGTAAAAGTACACATGATGACGAGATCGTGTATTACGCTGCAATTTGGTCTGGGG	1199
QY	443	yGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTy	463
Dd	1197	CGAAAAAGCCAGCTTGCGATTCGCTATTTCCCGTATCACCGAATTTGCCGACACGGGATA	125
QY	463	rThrProLeuLysValasnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSe	483
Dd	1257	CTCCCACTCTTTACGCCGAGCGGATCAGCGCATGACCGCGAGTGGGCAACTACAG	1316
QY	483	rPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerie	503
Dd	1317	CTTCACTCGGTCTTCAAGCCGGGATGAGTCCCTCTACCAAGCTGTGCGGCGCTA	1376
QY	503	uGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLysLysIleTr	523
Dd	1377	TGAGATCTTCACTCGGGGCGACATTCACAAAGATATCCCTACTGCGCTTTGGCTGTGA	1436
QY	523	pProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValProLe	543
Dd	1437	TGACGAATTCAGTCTGGTGGACCTAAGGATACGTGGCATATCAGAATATCCCTCTAT	1496
QY	543	u 543	
Dd	1497	T 1497	
RESULT 8			
AX534824			
LOCUS	Sequence 62 from Patent WO02068623.	1581 bp	DNA linear PAT 22-NOV-2002
DEFINITION	AX534824		
ACCESSION	AX534824		
VERSION	AX534824.1	GI:25261366	
KEYWORDS	Aspergillus niger		
SOURCE	Aspergillus niger		
ORGANISM	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;		
REFERENCE	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.		
AUTHORS	Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A., Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gustedt Von,W., Heinrich,O., Mater,D., Spreafico,F., Folkers,U., Hopper,S., Kemmer,W., Tan,P., Stiebler,J. and Albang,R.		
TITLE	Novel genes encoding novel proteolytic enzymes		
JOURNAL	Patent: WO 02068623-A 62 06-SEP-2002;		
FEATURES	DSM N.V. (

Alignment Scores:					
Pred. No.:	1.45e-62	Length:	1581		
Scores:	925.00	Matches:	221		
Percent Similarity:	51.46%	Conservative:	78		
Best Local Similarity:	38.04%	Mismatches:	192		
Query Match:	31.05%	Indels:	90		
DB:	6	Gaps:	15		
 US-09-712-338-2 (1-555) x AX534824 (1-1581)					
QY	4	TyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuProGlySerThr	23		
Db	4	TACTACTCTCTCGGGTTCGTGCCTTGTTGGGCC-----GGCGTGCCC-----	45		
QY	24	ProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGlyValIlysThrLeuThr	43		
Db	46	-----GTCTCCGGGGCAGTAGTTGTGGTTCGCCGCCACCGATCATCTCCACCACAG	96		
QY	44	ThraAlaAsnValThrIleArgTyrlLysGluProGlyAlaGluGlyValCysGluThr	63		
Db	97	GGATATCTGCATCCCGGCCCTCAACAACAGGTC---CCACCGGCATTGTGAGACT	153		
QY	64	ThrProGlyValIysSerTyrlValAspThrSerProGluSerHisThrPhe	83		
Db	154	GATCCCAAGTCTCAAGAGCTTCTCCGGTTACGTTCGATGCTGCTGAGCATGAGCACATCTTC	213		
QY	84	PheTrpPheGluAlaArgHis-----AsnProGluThrAlaProIleThrLeuTrpLeu	102		
Db	214	TTCTGGTTCTTCAGGCGCGCAACCAAGATATCCACCGAGGCTCCCTTGACCGTCTGGAIC	273		
QY	103	AsnGly-----GlyProGlySerAspSerLeuIleGlyLeuPheGluGlu	117		
Db	274	AATGGAGCATGCTGACCCCGTCTCTGGTTCCTCTCCATGATCGGCTTGTCCAAGAG	333		
QY	118	LeuGlyProCysHisValAsnSerThrPheAspAspTyrlIleAsnProHisSerTrpAsn	137		
Db	334	CACGGCCCATGGCCATATGACGCCAATGGCTCCGCTCTACAAACACCCCTACTCTCTGAAAC	393		
QY	138	GluValSerAsnLeuPheLeuSerGlyProLeuGlyValGlyPheSerTyrlSerAsp	157		
Db	394	AACCCAGCAACATGCTCTACATGACGACGCGCGTGCAGACCGGCTTCTCTACAGC---	450		
QY	158	ThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGly	177		
Db	451	-----ATTCCGGTCCCGGCTATGTGGATTCTTCCACAGACAATGGT	492		
QY	178	ValGlnGlyArgTyrlProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAla	197		
Db	493	TTTTATGGCGCATTTCTCT-----	510		
QY	198	GluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArg	217		
Db	511	-----CAG	513		
QY	218	ValGlnSerIysAspPheSerLeuTrpThrGluSerTyrlGlyHisTyrlGlyProAla	237		
Db	514	TACTCGCGGAACCTTCCACTTCCACCGAGAGATTATGGCGCCACCTACGGCCCCGTC	573		
QY	238	PheAsnHisPheTyrlGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyVal	257		
Db	574	TTTACAGAGTATCATCGAGGAGCAGAACCCCATCTCCAGCCGGGAGCC-----AAG	624		
QY	258	GlnLeuAsnPheAsnSerLeuIleIleAsnGlyIleIleAspGluAlaIleGlnAla	277		
Db	625	AGATCTCAACTGGGCGGTGTGATGATCGGCATGCTGTGATGACCCGATTATTCATPAC	684		
QY	278	ProTyrlTyrlProGluPheAlaVal-----AsnAsnThrTyrlGlyIleAlaValAsn	295		
Db	685	CAGGCGCTACTACAACTTTTACGGTATPATCCGGGCAACACATACGACTACTCCGCAATCAAC	744		
QY	296	GluThrValTyrlAsnTyrlMetIysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeu	315		
Db		-----	804		

TITLE Novel genes encoding novel proteolytic enzymes
JOURNAL Patent: WO 02068623-A 5 06-SEP-2002;
DSM N.V. (NL)

```

FEATURES
    source
        Location/Qualifiers
            1..2940
                /organism="Aspergillus niger"
                /mol_type="genomic DNA"
                /db_xref="taxon:5061"

```

BASE COUNT	691 a	-818 c	732 g	699 t
/db_xref="taxon:5061"				

Alignment Scores:	
Pred. No.:	1.79e-60
Score:	902.50
Percent Similarity:	49.45%
Best Local Similarity:	36.61%
Query Match:	30.30%
DB:	6
Length:	2940
Matches:	231
Conservative:	81
Mismatches:	206
Indels:	114
Gaps:	18

US-09-712-338-2 (1-555) x AX534767 (1-2940)

QY	4	Tyr	GlupheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuProGlySer	Thr	23
DB	456	TACTACTCTCTCTGGTGTCTGGTGGCC	-----GGCGTGGCC-----		497
QY	24	ProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGlyValLysThrLeu	Thr	43	
DB	498	-----GTCCTCCGGGCCAGATTGTGGCTCGGCCACGATCTCATTCACCAAG	548		
QY	44	ThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGlu	Thr	63	
DB	549	GGATATCTCGACATCCCGCTCGGTACAAAGTGC	---CCACCGGCATTTGTGAGACT	605	
QY	64	ThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThr	Phe	83	
DB	606	GATCCCAAGTGTCAAGAGCTTCTCCGGTACGTGCATGTGCTGAGCATGAGCACATCTC	665		
QY	84	PheTrpPhePheGluAlaArgHis	---AsnProGluThrAlaIleProLeuThrLeu	TrpLeu	102
DB	666	TTCTGGTCTTCGAGGCGGCAACCAAGATCCCAOCGAGGCTCCCTTGACCTGGATC	725		
QY	103	AsnGly	-----Gly	105	
DB	726	AATGG-AGGCATGCTGACCCCGGTCAATTATTTCTTCCAAATGCTAACCGTTCGTAGT	784		
QY	106	ProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsnSer	125		
DB	785	CCTGGTCTCTCTCCATCGATCGGCTTTTCCAAAGACAGCGGCCATGCGCATTCACGCC	844		
QY	126	ThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeu	145		
DB	845	AATGGCTCCGTCACAAACACCCTACTCTCTGGAACAACGCGACGAACATGCTCTACATC	904		
QY	146	SerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnPro	165		
DB	905	GACCAGCCCGTGCAGACCGGCTCTCTCTACAGC	-----ATTCCG	943	
QY	166	ValThrGlyValValGluAsnSerSer	-----	174	
DB	944	GTTCGGCTATGTGGATCTTCCACAGACAATGTTATTGCTGCTCCCTCCCGCCCTGC	1003		
QY	175	-----PheAlaGlyValGlnGlyArgTyrProThrIleAspAla	187		
DB	1004	CCCGACTATGCAGCGGATATGTTCTGTGGCACTTACTCC	---TACCCCAAGCTGAGCGCTT	1060	
QY	188	ThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPhe	207		
DB	1061	ACGGCTAATTCACCGACAAC	---GCGGCCCAACTTCTACCGGCCCTACAGGGTTTT	1117	
QY	208	LeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeuTrpThr	227		
DB	1118	ATGGGCGCATTTCT	-----CAGTACTCGCGCAACCTTCCATCTACCCACG	1165	

```

Db      2213 GAGCGGAGGCTGAAGCGGATGGGACAGCT 2243
RESULT 10
AX534809          AX534809          3080 bp      DNA      linear      PAT 22-NOV-2002
LOCUS             Sequence 47 from Patent WO02068623.
DEFINITION        AX534809
ACCESSION         AX534809.1 GI:25261332
VERSION           AX534809.1
KEYWORDS
SOURCE            Aspergillus niger
ORGANISM          Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE
AUTHORS           Edens, L., dijk Van, A.A., Krubasik, P., Albertmann, K., Stock, A.,
Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., guscedt von, W.,
Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,
Kemmer, W., Tan, P., Stiebler, J. and Albarg, R.
Novel genes encoding novel proteolytic enzymes
Patent: WO 02068623-A 47 06-SEP-2002;
DSM N.V. (NL)
FEATURES
Location/Qualifiers
1..3080
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT       755 a 803 c 735 g 787 t
ORIGIN
Alignment Scores:
Pred. No.:       1,04e-59      Length:      3080
Score:           893.00        Matches:     232
Percent Similarity: 47.93%      Conservative: 81
Best Local Similarity: 35.53%    Mismatches: 200
Query Match:     29.98%        Indels:     141
DB:              6            Gaps:       19
US-09-712-338-2 (1-555) x AX534809 (1-3080)
QY      4 TyrGluPheLeuSerValLeuProLeuVal-----AlaAlaSerTrpAlaLeuProGly 21
Db      490 TATCATGCTTCCAGCTTCTCCCGCATGTGAGTAGTCTGCTCTGTTGGGGGTCTTCGGG 549
QY      22 SerThrProAlaSerValGlyArg-GlnLeuProLysAsnProThrGlyValLysTh 41
Db      550 TCTAGGCAC-----CGCTCAATTCTCCCGAGCGCGAGGCATCACTGT 594
QY      41 rLeuThrThraLa-----AsnAsnValThrIleArgTyrLysGluProGlyAlaGlu-- 58
Db      595 GCTCAAGTCCAAGTGCATGAGAATGTGACTATTCTTTCAAAGAGGTGTGTGCAGA-GT 653
QY      59 -----GlyValCysGluTh 63
Db      654 ATCTAGAAATAGCTTTATGCTCGATGCCCGCTGCTGATGTGACGCTGGAAATTTGGGAAC 713
QY      63 rThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro----- 78
Db      714 TAGCCCGGTGTCCGATCTTATTCGGGCTATGTATACACCTTCCCGCGGTTTCTCTTCCGA 773
QY      78 ----- 78
Db      774 CGGACAGGAGAGTGCAGGATTATCTATCAACACGTAAGCCCAATCTCGAAACATTGGA 833
QY      79 -----Glu-SerHisThrPheThrPhePheGluAlaArgHisAsn 93
Db      834 GGATAGCAATTAATCTAGACCTCAACAGCTTCTTTGGTTTTCAGACCCGCAAGATC 893
QY      93 roGluThrAlaProIleThrLeuTrpLeuAsnGlyProGlySerAspSerLeuIleG 113
Db      894 CCAGCAATGCGCTCTGGCCATCTGGCTCAATGCGGTCCGGGTGGCTCGCTCATGG 953
QY      113 lyLeuPheGluGluLeuGlyProCysHisValAsnSer---ThrPheAspAspTyrIleA 132

```

```

Db      1968  TTCTGGATGCAATTGGCTTACTCTCTCGACAGTGTGTAAAGTACACATGATGATGACGA 2027
Qy      433  AspaAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyr 452
Db      2028  GATCGTATTAGCCCTGCAATTGGTGGGCGGAAAGCCAGCCTTGCAGTTCGGTAT 2087
Qy      453  SerGlnAlaAlaGlnPheArgSerAlaGlyTyrTrpProLeuLysValAsnGlyValGlu 472
Db      2088  TCCCGTATCACCGAATTTGCCGACACGCGGATATCCCCACTCTTACGCCCGCGGATC 2147
Qy      473  TyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHis 492
Db      2148  AGCGCATGACCCCGGCTGCGGCAACTACAGCTTCACTCGCTTCTCCAAAGCCGGCAT 2207
Qy      493  GluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGly 512
Db      2208  GAGTCCCTCTTACCAGCTGTGCGGCGTATGAGATCTTCATCGCGCGGCGACATTTCAAC 2267
Qy      513  TrpAspIleAlaGluGlyGlnLysIleTrpProSerTyrLysThrAsnGlyThrAla 532
Db      2268  AAGATATATCCCTACTGCGCTCTTGGCTGTGATGACGAATCCAGTCTGGTTGGACCTAAG 2327
Qy      533  ThrAlaThrHisThrGlnSerSerValProLeu 543
Db      2328  GATAGTGGGATATCAAGATATCCCTCTATT 2360

```

```

RESULT 11
LOCUS   D86560
DEFINITION Schizosaccharomyces pombe cpyl gene for carboxypeptidase Y,
complete cds.
ACCESSION D86560.1
VERSION   GI:3046860
KEYWORDS cpyl; carboxypeptidase Y.
SOURCE    Schizosaccharomyces pombe (fission yeast)
ORGANISM  Schizosaccharomyces pombe
          Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
          Schizosaccharomycetales; Schizosaccharomycetaceae;
          Schizosaccharomycetes.
REFERENCE 1 (sites)
AUTHORS  Tabuchi,M., Iwahara,O., Ohtani,Y., Ohuchi,N., Sakurai,J.,
          Morita,T., Iwahara,S. and Takegawa,K.
TITLE     Vacuolar protein sorting in fission yeast: cloning, biosynthesis,
          transport, and processing of carboxypeptidase Y from
          Schizosaccharomyces pombe
JOURNAL  J. Bacteriol. 179 (13), 4179-4189 (1997)
MEDLINE  97352672
PUBMED   9209031
REFERENCE 2 (sites)
AUTHORS  Takegawa,K., Tabuchi,M., Iwahara,O., Morita,T. and Iwahara,S.
TITLE     Cloning and characterization of cpyl gene from Schizosaccharomycetes
          pombe
JOURNAL  Unpublished
REFERENCE 3 (bases 1 to 4308)
AUTHORS  Takegawa,K.
TITLE     Direct Submission
JOURNAL  Submitted (16-JUL-1996) Kaoru Takegawa, Kagawa University,
          Department of Bioresource Science, Faculty of Agriculture, Ikenobe
          2393, Kita-gun, Miki-cho, Kagawa 761-07, Japan
          (E-mail:takegawa@ag.kagawa-u.ac.jp, Tel:087-891-3116,
          Fax:087-898-7295)
COMMENT  On Apr 15, 1998 this sequence version replaced gi:2274843.
FEATURES
source   Location/Qualifiers
         1..4308
            /organism="Schizosaccharomycetes pombe"
            /mol_type="genomic DNA"
            /db_xref="taxon:4896"
            812..3820
               /gene="cpyl"
            812..3820
               /gene="cpyl1"
            /codon_start=1

```

```

/product="carboxypeptidase Y"
/protein_id="BAA25568.1"
/db_xref="GI:3046861"
/translation="MLMKQTFLLTTCVSAQFNQYVBPBQNGGDIIVPKDFYKRG
EDFTRQEESSAPLMNPVPRDEAEAPHPKGGHEFNDDEDDTALEHPGKDKLDSF
LOPARDLHTYSDDLNIIDDEDEHVREKPHDSADAPRRKHGKCGKGGHKGX
HAKGKSKSPKPEDDSVFDDERPKHGFDDREDFPAHHEPGEHMPFPMHHPKGE
HMPFPMHHPGEHMPFPMHHPGEHMPFPMHHPGEHMPFPMHHPGEHMPFPMHHP
EHMPFPMHHPGEHMPFPMHHPGEHMPFPMHHPGEHMPFPMHHPGEHMPFPMHHP
KPKDKKHGKPKDKKHGKPKDKKHGKPKDKKHGKPKDKKHGKPKDKKHGKPKDKKH
NSVPEVETAEESFSLDTLGLTYGYNIPVHIOENAPKDSSTPLFEFDDDLLESLTP
EQPAILMLKRAEGIDPMTRDQHPKPSNAQPADSRPVAFVSQENGSHVAKAF
PDHLTRVKDSKESLGDITVKQTYGLDVEDDRHLFFWFFESRNDPNDPVVLWNGG
PGSSLTGLFMELGPSSINIEFLKPNPHSNWSNAGVIFLDQPIINTGFSNGDVLG
TVTAKGVYAFNLFFAFQYALHDPHIAGESYAGVIFPOFAKIMEHNGANFEVA
SGYEMEXOYINLKSVLINGTLPLVOYFYFKVACSPGPIQMSQECRITGAYDT
CAKILTCYOTGTPTVCIGASLYCNAMIGPFTKGLNIDIRECDRQDEHLCPETG
AIESYLNQEFQOALGVYDKNTEVNIIGLFGDWMRKTFEDDDVLAELGLVL
IYAGDADYINYMGNENATDALEWAGREFYEALKPWSNPGKAGRKSPKNTFYLR
LYEAGHMVFPNQPEASLEMLNSWTDGSLFA"
BASE COUNT 1184 a 958 c 871 g 1295 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 7.48e-41 Length: 4308
Score: 652.50 Matches: 170
Percent Similarity: 48.47% Conservative: 67
Best Local Similarity: 34.76% Mismatches: 168
Query Match: 21.90% Indels: 85
DB: 8 Gaps: 15

US-09-712-338-2 (1-555) x D86560 (1-4308)
Qy      49  ThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyValLys 68
Db      2519  ACTTTCGCGTTAAGATTCCAACTGAATCACTTGTGTATTGACACT-----GTTAAG 2572
Qy      69  SerTyrSerGlyTyrValAspThrSerProGlySerHisThrPhePheThrPhePheGlu 88
Db      2573  CAATACACCGGTTACTTAGATGCGAAGATGACAGACATCTTTCTCTGGTCTTTTGAA 2632
Qy      89  AlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySer 108
Db      2633  TCTAGAAATGATCCGAGATGATCCGCTGTGTGTGTTGAACGGTGTCTCTGGTTCG 2692
Qy      109  AspSerLeuIleGlyLeuPheGluGlyProCysHisValAsn---SerThrPhe 127
Db      2693  TCTTCCCTTACTGGTTTGTTCATGGAAATTAGGTCCTTCTTCAATCAACATGAGACTCTT 2752
Qy      128  AspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGln 147
Db      2753  AAACCCGAATATAACCTCACAGTTCCAACTCCAACTCCATCTTATCTTTTGGATCAA 2812
Qy      148  ProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThr 167
Db      2813  CCTATCAACACGGGTTTCAGCAACGGAGAT-----GACTCGGTTCTTGACACTGTACG 2866
Qy      168  GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAla 187
Db      2866  -----
Qy      188  ThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPhe 207
Db      2867  -----GCTGTAAGGATGTTTATGCG-ATTCCTGAACCTTTT 2901
Qy      208  LeuSerGlyLeuProSerLeuAspSerArgValGlnSer-LysAspPheSerLeuTrpThr 227
Db      2902  CTTTCCCAAGTTCCCTCA-----GTAACGTCATTTGGACTTTTCACATTCCTGG 2949
Qy      227  rGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnGln 247
Db      2950  TGAATCCTATGCTGCGCCATTACATCCCCAGTTTTCGAAGGAAATATATGGAGATAACCA 3009

```

trehalose-phosphatase; tricarboxylate transport; tRNA-Leu;
ureidoglycolate hydrolase.
SOURCE Schizosaccharomyces pombe (fission yeast)
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
1 (bases 1 to 37000)
Oliver K., Harris, D., Wood, V., Barrell, B.G. and Rajandream, M.A.
Direct Submission
Sequencing (30-JUN-1997) Schizosaccharomycetes pombe chromosome I
submitting project, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk
On Feb 13, 1998 this sequence version replaced gi:2239193.
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Spsplice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct/
splice donor/acceptor sites.
CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid c19G12 is overlapped at the 3' end by cosmid c23A1.
FEATURES
source
Location/Qualifiers
1..37000
/organism="Schizosaccharomyces pombe"
/mol_type="genomic DNA"
/strain="972h"
/db_xref="taxon:4896"
/chromosome="I"
/map="IR"
/clon="cosmid c19G12"
complement(join(1..823,869..926,1018..1140,1183..1402,
1451..1688))
/gene="SPAC19G12.01c"
/note="synonym: cut20"
complement(join(1..823,869..926,1018..1140,1183..1402,
1451..1688))
/partial
/gene="SPAC19G12.01c"
/note="SPAC19G12.01c, len:487"
/codon_start=1
/label=cut20+
/product="cyclosome/apc subunit cut20/apc4"
/protein_id="CAB53521.1"
/db_xref="GI:5791550"
/translations="MVSQKPKPKKFEWNETNRVERVSGLKRVLICPSMELIAI
LTCSHLIIICRSQRIWDVDFHDLATELCWNHGNLIIVGFKNGELKIIIDSGTEL
VEQPSRDLAVIMTWAMQETIVNEKRNDFIDATYMLGLTLPSSAKERIFSSK
AIAQFPRPERGEGNNKKVELLSILDERGIRYINFMSSYKIGSDLSKALNLGVPI
HSITNDIAHYVLICKGTNISLKYMLKNDLSIVDIATWTRQHLVRYLKEVIN
NMTVEFDNVTKGEASIKTFIDALVSKYSITTFPSQLELYQFINNGIPSDLLKWTIL
ERVGDRLVKWERMVNSYLSLIIFCOEVPACERITVLLSSARKSIWGHMKGTIL
LDAKLVEDLATYLGNNVFSNCLTFEKKYMKHFIISWNLVAILVEFNITSEFPSSIP
QEI TEHNTEWIVYRHSLSKTFYFMGFKPLQLRD"
complement(join(2169..2404,2446..3702,3774..4239,
4342..4767))
/gene="SPAC19G12.02c"
/note="synonym: pms1"
complement(join(2169..2404,2446..3702,3774..4239,
4342..4767))
gene
CDS
gene
CDS

```

4342..4767))
/gene="SPAC19G12.02c"
/note="SPAC19G12.02c, len:794"
/codon_start=1
/product="DNA mismatch repair protein pms1"
/protein_id="CAB10113.1"
/db_xref="GI:2239195"
/db_xref="SWISS-PROT:P54280"
/translation="MSTVKPIDANTVHKISGQVITDVASAKVELVNSLDGATTIE
IRFNGINSLEVNDGSDIGADYESIIGKHFSTKIDPDLBALQTFGRGALSS
LCAGVOVISTATQNEAPKGVQLNDHEGSKOKLITPFGSTVSWNDLFTLPVR
KLEKNYKRFSSKALSULQAYATITSTKRFVWYHQTNSKGLQLSTNSKMDKINIM
NVFQTVSSSLIPNDGIIIEYISRPVSGSTRASNERQMLFINRLVNLPKIARVIOE
VFQYSMAQSPFFAINRITNGTIDINVPDKKSVFLSEDSIIIEFTKSLONLCESC
CHAIKSCGSIQSPDSGSDTDQLPQSIIPATESITSDPSFSYKSPCKRK
LVEATQAPATISTVAEGASLAQVSKPIPERLQDKMRSSPLNEKVTFASERMKKLA
LFASSTTSQKTKTIDSSFPPLKQPKINKESSPNLLNDPSPASTVAKTINLNEIESV
HNBSVSTLSIPKEQTSVANRIPSKTAALQKLKFOQSLDGLNKFSSKINISISG
VQNDIVRSALLKFSNGKGVVHDSIDENQBDHLNLTVKADFLMRVVGQFNRGFIIV
VQGNLFIDQHASDERFNEHLKSNLVINSQDLVLPKRLDLAATEVTILIDHDLIR
RKGFVAIDNQRVGNRTLLSVPTSKNVIPTSDLEIISVLSSEHPQIDPFSRLER
MLKAGCRSSVMIGRALTISEMNTIVRHLAELSKPWCNCPHGRPTMRHLRLKDI"
complement(2405..2417)
/misc_feature
/gene="SPAC19G12.02c"
/note="taactctctgag, splice branch and acceptor"
complement(2440..2445)
/misc_feature
/gene="SPAC19G12.02c"
/note="gtaagt, splice donor sequence"
complement(join(3623..3702,3774..4239,4342..4758))
/misc_feature
/gene="SPAC19G12.02c"
/note="Match to PF0119 DNA_mis_repair, DNA mismatch
repair protein score 529.60"
complement(3703..3719)
/misc_feature
/gene="SPAC19G12.02c"
/note="craacaataatttag, splice branch and acceptor"
complement(3768..3773)
/misc_feature
/gene="SPAC19G12.02c"
/note="gtagtg, splice donor sequence"
complement(4240..4253)
/misc_feature
/gene="SPAC19G12.02c"
/note="ctaactgaatgag, splice branch and acceptor"
complement(4336..4341)
/misc_feature
/gene="SPAC19G12.02c"
/note="gtaagt, splice donor sequence"
complement(4468..4488)
/misc_feature
/gene="SPAC19G12.02c"
/note="PS00058 DNA mismatch repair proteins mutL"
5410..6372
/gene="SPAC19G12.03"
5410..6372
/gene="SPAC19G12.03"
/note="SPAC19G12.03, len:320, LOW SIMILARITY:Yersinia
pseudotuberculosis, CAB46604, hypothetical 33.7 kd
protein., (293 aa), fasta scores: opt: 155, E():0.003,
(27.5% identity in 265 aa)"
/codon_start=1
/label="SPAC19G12.03"
/product="possible polysaccharide deacetylase"
/protein_id="CAB10114.1"
/db_xref="GI:2239196"
/db_xref="SPTRMBL:O13842"
/translation="MYETRDLTGNAGKPVDTNPWNNNSKIASFVNVYEGGERSILY
EDEGETFLTEAGLPFNPVRERSIESCFYGRGCGPWRILNLFKHKVPTQWAI
GQAVENPVVVGAMEAGCEGVGSHRWINVEGVPETIEYEHKKSQVQAIQASPSNS
APRSYNTGPASINTKLVQCVKDLGLPQFPDSFYDSDYNDIPYVADPLASKPGEADDK
GLLIIVPTLENDMKYAVAPGFCNSDDFYTYARDAPDVLVEGLEGAPKQMTIGLHCR
LTGRGPRFGLQKLEHITSKGVVWVATREQIAQASAKHPYKA"
5608..6042
/misc_feature
/gene="SPAC19G12.03"
/note="Pfam match to entry PF01522 Polysac_deacet"
6888..7463
/gene="SPAC19G12.04"
6888..7463

```

```

/gene="SPAC19G12.04"
/note="SPAC19G12.04, len:191, SIMILARITY:Saccharomyces
cerevisiae, DAL3 YEAST, ureidoglycolate hydrolase, (195
aa), fasta scores: opt: 360, E():5e-17, (36.5% identity in
200 aa)"
/codon_start=1
/label="SPAC19G12.04"
/product="putative ureidoglycolate hydrolase"
/protein_id="CAB10115.1"
/db_xref="GI:2239197"
/db_xref="SPTRMBL:O13843"
/translation="MEAKKLYAALTDIEAPFGSVVQQDDVVMVANGGTAKKYLK
VRSIQYKSSASRKGWNFSTHPSVHPANDEHAFQISVLEHPTTQTFIFM
CRSDQAYLIIVAPNAPQMPDQNTQAFVARGAQQGVTSAGVWHAPMVTIGKETML
AARNYENGVAEDDCQVQSTESPSEIEVFIKIST"
8397..9272
/gene="SPAC19G12.05"
8397..9272
/gene="SPAC19G12.05"
/note="SPAC19G12.05, len:291, SIMILARITY:Saccharomyces
cerevisiae, TXTP YEAST, tricarboxylate transport protein,
(299 aa), fasta scores: opt: 738, E():0, (49.0% identity
in 290 aa)"
/codon_start=1
/label="SPAC19G12.05"
/product="putative tricarboxylate transport protein"
/protein_id="CAB10116.1"
/db_xref="GI:2239198"
/db_xref="SPTRMBL:O13844"
/translation="MSTVAKTNKSNKPGPVKSIAGGVAGAIEISITVPABFAKTR
LQLYRVNVEGKAKLPFGLEWYKSGSTVIVNSLKAARFFAFDSIKKSLSDSHGLT
GRTVLIALGAGVAESVLVLTPEESIKTAIIDDKRPNRKLKGLQASRIIVHENGIR
GLYRLAATVARGAANSQVRFATYNSIKQSLQSLRPDEKLSTVTFVLSGVAGIIV
YCTQPIDTVKSRMOSLSASKEYKNSHCAYKILTDQGLLRFLWSGATPRLARLILSGEI
VFTVYKVMELKPF"
8397..9020
/misc_feature
/gene="SPAC19G12.05"
/note="PFam match to entry mito_carr PF00153,
Mitochondrial carrier proteins"
8505..8531
/misc_feature
/gene="SPAC19G12.05"
/note="PS00215 Mitochondrial energy transfer proteins
signature"
complement(10121..10516)

```

```

Alignment Scores:
Pred. No.:      Length:      37000
Score:          652.50
Percent Similarity: 48.47%
Best Local Similarity: 34.76%
Query Match:      21.90%
DB:               8
Gaps:             15

US-09-712-338-2 (1-555) x SPAC19G12 (1-37000)
QY 49 ThrIleArgTyrLysGluProGlyAlaGluValCysGluThrThrProGlyVallys 68
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 20080 ACTCTTCGGTTAAAGATTCCAACCTGAATCATTGGTATTGACACT-----GTTAG 20027

QY 69 SerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheThrPheGlu 88
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 20026 CAATCACCGGTATTAGATTCGAAGATGACACACTTTTCTCTGTTCTTGA 19967

QY 89 AlaArgHisAsnProGluThrAlaProIleThrLeuTlpLeuAsnGlyGlyProGlySer 108
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19966 TCTAGAAATGATCCGAGAAATCATCCCGTCGTGTGTGTGTGAACGGTGTCTGTTGC 19907

QY 109 AspSerLeulleGlyLeuPheGluGluLeuGlyProCysHisValAsn---SerThrPhe 127
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19906 TCTTCCCTTACTGGTTTCTTCATGGAATTAGTCTCTTCTCAATCAACATTGAGACTCTT 19847

QY 128 AspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGln 147
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

19846 AAACCCGAATATAACCTCCTCACAGTTGGAACTCCAATGCTTCAGTTATCTTTTGGATCAA 19787
148 ProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThr 167
19786 CCTATCAACACGGGTTTCAGCAACGGAGAT-----GACTCGGTCTTGACACTGTACG 19733
168 GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAla 187
19733 ----- 19733
188 ThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPhe 207
19732 -----GCTGTAAGGATGTTTATGC-ATTCCTTGAACCTTTT 19698
208 LeuSerGlyLeuProSerLeuAspSerArgValGlnSer-LysAspPheSerLeuTrpThr 227
19697 CTTTGGCAAGTTCCTCA-----GTACGCTCATTTGGACTTTCACATGCTGG 19650
227 rGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnG1 247
19649 TGAATCCTATGCTGGCATTACATCCCCAGTTTCCCAAGGAAATTTATGGAGCATAACCA 19590
247 uArg-----IleAlaAsnGly-----SerValAsnGlyValGlnLeuAsnPheAs 262
19589 AGGTGCTAACTTCTTTGTCGACGGCTTATGAAATGGAGAGCAATACATCAATTTGAA 19530
262 nSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProG1 282
19529 GAGTGTCTGATGGAAATGGTTTGTACTGATCCTTTGGTCCATACTACTTTTACGGAAA 19470
282 uPheAlaValAsnAsnThrTyrGly-----IleIle 292
19469 AATGGCTTCGAGAGCCCTTAAGTCTTATATGTCCTCCCAAGAGGAATGTGATGCATTAC 19410
292 salAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCy 312
19409 TGGTGCTATGATACC----- 19392
312 sGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLe 332
19391 CGCTAAGCTAATCACTGGCTGTATACAGACTGCTTTACT-----CCTGT 19347
332 uCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAl 352
19346 TTGCATTGGTGGCTCTTGTATTGCAATAACGCTATGATGGACCA---TTTACTAAGAC 19290
352 aGlyArgGlyValTyrAspIleArgHisProTyrAspAspProThrProSerTyr-- 371
19289 TGGACTCAACATTATGATATTCGTGAAGATCCCGTGACCAAGAGCACTTATGCTACCC 19230
372 -----TyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyVa 387
19229 CGAAACCGGTGCAATTGAGAGTTACTTTGACCAAGAATTTGTTCAAGAAGCTTTGGGAGT 19170
387 lAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAs 407
19169 TGAATACGATTACAAGGGATGCAATACCTGAAGTAACCAATTTGCTTTTCAAGGGTGA 19110
407 pPheValTrpProAsnPheIleGluAspLeuGluIleLeu-----AlaLeuProVa 425
19109 CTGGATCGGTAAGACTTTCGTGACGAATGACCGCAATCTTAGAAGCTGGCTTACCCGT 19050
425 lArgValSerLeuIleTyr---GlyAspAlaAspTyrIleCysAsnTrpPheGlyGly 444
19049 T-----CTTATCTATCGCGGTGATGCTGACTACATTTGCAATTACATGCGCAATGA 18999
444 nAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGluPheArgSerAlaGlyTyrTh 464
18998 AGCTTGACCGACCACTTGTAGTGGGCTGGTCAACGTGAGTTTATGAGGCCGCAATGAA 18939
464 rProLeuLysValAsnGlyValGluTyrGluThrArgGluTyrGlyAsnPheSerPh 484
18938 GCCTTGGAGTCTTATGGAAGGAGCTGGTCTGTTAGTCTTTCAAAACCTTTGTTA 18879

```

484 eThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuG1 504
18878 TCTTCGCTCTTACCAAGCTGGACACATGGTTCCTTTCAACCAACCGAAGTAGTTAGTA 18819
504 nLeuPheAsnArgThrIlePheGly 512
18818 AATGTTGAACAGCTGGATAGATGGT 18794

RESULT 13
AX594884 1527 bp DNA linear PAT 14-FEB-2003
LOCUS Sequence 538 from Patent EP1258494.
DEFINITION AX594884
ACCESSION AX594884
VERSION AX594884.1 GI:28396526
KEYWORDS
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1
AUTHORS Bauer, A., Gavin, A.C., Grandi, P., Krause, R., Kruse, U., Kuster, B.,
Marzioch, M., Schultz, J. and Superti-Furga, G.
TITLE Multiprotein complexes from eukaryotes
JOURNAL Patent: EP 1258494-A 538 20-NOV-2002;
CELLZONE AG (DE)
FEATURES
Location/Qualifiers
1..1527
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
BASE COUNT 439 a 300 c 319 g 469 t
ORIGIN

Alignment Scores:
Pred. No.: 3.82e-40 Length: 1527
Score: 636.00 Matches: 167
Percent Similarity: 46.20% Conservative: 70
Best Local Similarity: 32.62% Mismatches: 137
Query Match: 21.35% Indels: 78
DB: 6 Gaps: 14

US-09-712-338-2 (1-555) x AX594884 (1-1527)
QY 19 LeuProGlySerThrProAlaSerValGlyArgArgGlnLeu-----ProLys 34
Db 127 CTTCCACAGATACACAGCAAAACCCCTTAAATTTGGACCGTTTGAATCAGATATCCGCTG 186
QY 35 AsnProThrGlyValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGlu 54
Db 187 TTTACAACTTTATTCTTCTCTGTGGACACAGATTACAGTTTGAGACTTAGAACAGTAGAT 246
QY 55 ProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal 74
Db 247 CCTTCTAACTAGTAGAATT-----GACACCGTAAACAATAGTTCGGGTACATG 294
QY 75 AspThrSerProGluSerHisThrPheThrPhePheGluAlaArgHisAsnProGlu 94
Db 295 GACTATAGGAATTCGAACACACTTTTCTTGTGTTTGTAAAGTAGAAGACGATCCCTGCT 354
QY 95 ThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 114
Db 355 AAGACCACTATTCTTTGTTAATGTTGGACCTGTTGTTCTTCCTGTTTACTGGGTG 414
QY 115 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis 134
Db 415 CTATTTGAACATAGGCCCTCATCAATTGGCGCGATATGAACCAATCCCAATCCCTAT 474
QY 135 SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer 154
Db 475 TCTTGGAAATTAACGGTCTTCAATCATCTCTTTAGACAGCACTCGGAGTCGGTTTCC 534
QY 155 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer 174

535 TATGCTGAT----- 543

175 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsn 194

544 -----GAAAAGTCTCTCTACAAA 564

195 LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu 214

565 TTAGCAGGCAAGATGCGCTACATTTCTCTGGAAATGTTTTGAAGCTTTTCTCATTTA 624

215 AspSerArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyr 234

625 -----CGCTCCAAACGATTTCCACATTCAGCGCAATCTCTATGCGAGCAATAT 672

235 GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer 253

673 ATCCCTCAAAATGACATGAGATCGTTGTCAGAACCCCTGAAAGA----- 717

254 ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGlu 273

718 -----ACGTTCAAATTTAACTTCAGTTATGATTGGTAAATGGTATCACAGACCT 765

274 AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys 292

766 TTGATTCAGCAGATTAATATGAACCAATGGCATCGGGAAAGGGGGCTATCACCGCTGT 825

293 AlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 312

826 CTCTCATCAGAGAAGATGAGAAATAGTAAGCTGCAGGCTCGT----- 870

313 GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu 332

871 -----TGTCGTAGGTTGAACAAGTTATGTTATGCTTCTTAATCAAGT 912

333 -----CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAla 350

913 TTACATGCATAGTCGCCACCTGCTTACGTGCACTCTGCACATTTTGGAAACCGTACATTAA 972

351 PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro----- 366

973 --ACAGGACTCAACGCTCTATGACATTAGACGGCCCTGTGAAGATAATAGTACTGATGGT 1029

367 -----ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 383

1030 ATGTGTTATACAGGCTCCGCTATGTGACCCAGTATATGAATTTTCTGAAGTTCAGAA 1089

384 AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPhe 402

1090 ACGTACGGTCCGACGTGCATAATATTCTGGCTGTGATATGATGACGTGTTCCACCGGATTT 1149

403 GlnGlnThrGlyAspPheValTyrProAsnPheIleGluAspLeuGluIleLeuAla 422

1150 TTGTTTACGGCGATGGAAGTAAACCA---TTTCAACAATATATTGCTGAATTTAAAT 1206

423 LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGly 442

1207 CACAACATTCGGGTATTAAATATATATGCGGTGATAAGGATTATATTTGTAATTGGCTGGGA 1266

443 GlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheAsnArgSerAlaGly 462

1267 AACCATGCTTGGTCCAAATGAGTTGGAATGATCAATAAACGTAGGTATCAGAGAAGGATG 1326

463 TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 480

1327 TTAAGACCATGGGTCACTAAAGAAACAGGTGAAGAGTTGGGACCAAGTCAAGAACTATGGC 1386

481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 500

1387 CCTTTCACCTTTTGAGAATATACGATCCGGTCATATGTTGCCCTATGATCAACCGGAG 1446

501 AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 512

Db 1447 GCAAGTTTGAAATGGTCAACAGTTGGATTTCGGT 1482

RESULT 14
SCVB139W

LOCUS SCVB139W 2016 bp DNA linear PLN 11-AUG-1997
S.cerevisiae chromosome II reading frame ORF YBR139W.
Z36008 Y13134
ACCESSION Z36008.1 GI:536435
VERSION
KEYWORDS
SOURCE
ORGANISM
Saccharomyces cerevisiae (baker's yeast)
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
1
Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Baclet,M.C.,
Barthe,C., Baur,A., Becam,A.M., Biteau,N., Boles,E., Brandt,T.,
Brendel,M., Brueckner,M., Bussereau,F., Christiansen,C.,
Contreras,K., Crouzet,M., Cziepluch,C., Demolis,N., Delaveau,T.,
Daignon,F., Dondet,H., Duesterhus,S., Dubois,E., Dujon,B., El
Bakkoury,M., Etian,K.D., Feuerhahn,M., Fiers,W., Fobo,G.M.,
Fritz,C., Gassenhuber,H., Glansdorff,N., Goffeau,A., Grivelli,L.A.,
de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K.,
Jacq,C., Jacquet,M., Jauniaux,J.C., Jonniaux,J.L., Kallioe,T.,
Kiesau,P., Kirchdrath,B., Koetter,P., Korol,S., Liebl,S., Mallet,L.,
Lohan,A.D.E., Louis,E.J., Li,Z.Y., Maat,M.J., Mallat,L.,
Mannhaupt,G., Messenguy,F., Miosga,T., Molemans,F., Mueller,S.,
Nasr,F., Obermaier,B., Perea,J., Pierard,A., Piravandi,E.,
Pohl,F.M., Pohl,T.M., Potier,S., Proft,M., Purnelle,B., Ramezani
Scherens,B., Schwarze,C., Skala,J., Slonimski,P.P.,
Smits,P.H.M., Souciet,J.L., Steensma,H.Y., Stucka,R.,
Urrutazu,A., van der Aart,Q.J., van Dyck,L., Vassarotti,A.,
Vetter,I., Vierdeels,F., Visser,M., Wagner,G., de Wergifosse,P.,
Wolfe,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Kleine,K.
Complete DNA sequence of yeast chromosome II
EMBO J. 13 (24), 5795-5803 (1994)
95112788
7813418
2 (bases 1 to 2016)
Becam,A.M., Herbert,C.J., Nasr,F., Slonimski,P.P. and Zagulski,M.
Unpublished
3 (bases 1 to 2016)
MIPS.
Direct Submission
Submitted (30-AUG-1994) Data collected by MIPS on behalf of the
European yeast chromosome II sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 19a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
FEATURES
Location/Qualifiers
1..2016
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="II"
328..1854
/note="ORF YBR139W"
/codon_start=1
/protein_id="CAA85097.1"
/db_xref="GI:536436"
/db_xref="SWISS-PROT:P38109"
/translation="WKYLNLFVQLLISIKYVGFGRAPSLFEDDTTFANLDKQKLP
QNTQTLKLRHNDPDLPTTFISVSDTDSLRLRTVDPKSLGDTVQKSGYMDYKD
SKHFFYFFERGNDFANDPIILWLGNGPGCSFTGLLPGSSIGADMKPILHNPYSW
NNASMIILEOPFLGVFGYSGEKSVSTKLAKDAYITLFLPEAFHLRGNDFHAGE
SYAGHIYIQAIAHEIWKNPENFTLTSWINGNITDPLIQADYVEPWACCKGGYHPVL
SSECEKMSKAGRCRKNLKYASKSLPCLIVATAYCDLSALLEPYINTGLNVYDIRG
PCEDNSTGMCYTGRLYVDQWMPFVEQVETLGSVDHNSGCDNDVFTEFLTGDSKSP
FQQYIAELNHLIPVLIYAGDKQICNWLGNHWSNELEWINKERYQRRLRPWVSKE
TGEELQGVKNYGPFTFLRIYDAGHMVYDQPEASLEMVNSWISGNRAFSDLSTLENAS
"

BASE COUNT 608 a 379 c 392 g 637 t
ORIGIN

Alignment Scores:

Pred. No.:	5,44e-40	Length:	2016
Score:	636.00	Matches:	167
Percent Similarity:	46.29%	Conservative:	70
Best Local Similarity:	32.62%	Mismatches:	197
Query Match:	21.35%	Indels:	78
DB:	8	Gaps:	14

US-09-712-338-2 (1-555) x SCYBRI39W (1-2016)

QY	19	LeuProGlySerThrProAlaSerValGlyArgArgGlnLeu	-----ProLys 34
DB	454	CTTCCACAGATACACAGCAAAACCTTTAAATTTGGACCGTTTGAATCAGATGATCCGCTG 513	
QY	35	AsnProThrGlyValLysThrLeuThrAlaAsnAsnValThrIleArgTyrLysGlu 54	
DB	514	TTTCAACCTTTATTTCTCTCTGTCGACACAGATTACAGTTTGAGCTTAGACAGTAGAT 573	
QY	55	ProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal 74	
DB	574	CCTTCTAAACTAGGAATT-----GACACCGTAAACAAATGGTTCGGGTTTACATG 621	
QY	75	AspThrSerProGluSerHisThrPheThrProPheGluAlaArgHisAsnProGlu 94	
DB	622	GACTATAAGGATTCACAAACACTTTTCTACTGGTTTCTTGAAGTAGGAACATCCTGCT 681	
QY	95	ThrAlaProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 114	
DB	682	AACGACCCCAATATTCTTTGGTTAAATGTGACCTGGTTGTTCTCTGTTTACTGGTTG 741	
QY	115	PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis 134	
DB	742	CTATTGGAAGTGGCCCTCATCAATGGCGCGATATGAAACCAATCCACAAATCCCTAT 801	
QY	135	SerThrAsnGluValSerAsnLeuPheLeuSerGlnProLeuGlyValGlyPheSer 154	
DB	802	TCTTGAATTAATACCTTCATGATCTCTTAGAACACGCCATCGAGTCGCGCTTTTCC 861	
QY	155	TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer 174	
DB	862	TATGGTGTAT-----870	
QY	175	PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsn 194	
DB	871	-----GAAAAAGTCTCTCTACAAAA 891	
QY	195	LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu 214	
DB	892	TTAGCAGGCAAGATCGTACATTTCTCGGAATGTTTTTGAAGCTTTTCTCTCATTTA 951	
QY	215	AspSerArgValGlnSerLysAspPheSerLeuThrThrGluSerTyrGlyHisTyr 234	
DB	952	-----CGCTCCACGATTTCCACATTCGAGCGGAATCCTATGCAGACATTAT 999	
QY	235	GlyProAlaPhePheAsnHisPheTyrGluAlaAsn---GluArgIleAlaAsnGlySer 253	
DB	1000	ATCCCTCAAAATTCACATGATGATCGTTGTCAAGAACCTCAAGAA-----1044	
QY	254	ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGlu 273	
DB	1045	-----ACGTTCAATTTAACTTCAGTTATGTTGGTATGTTATGTTATCAGACCT 1092	
QY	274	AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys 292	
DB	1093	TTGATTCACGACGATTATTATGAACCAATGCATGCGGAAAGGGGGCTATCACCCTGTT 1152	
QY	293	AlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 312	
DB	1153	CTCTCATCAGAGAAATGTGAAAAATGATTAAGTAACTCGAGTCGT-----1197	
QY	313	GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu 332	
DB	1198	-----TGTCGTAGGTGACAAAGTTATGTTATGTTCTTAATCAAGT 1239	

QY	333	-----CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAla 350	
DB	1240	TTACCATGCATAGTCGCCACCTGCTTACTGTGACTCTGCATCTTGGAAACCCGTACATTAA 1299	
QY	351	PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro-----366	
DB	1300	---ACAGGACTCAACGCTCTGTACATTAAGAGGCCCTGTGAAGATAAATAGTACTGATGGT 1356	
QY	367	-----ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 383	
DB	1357	ATGTGTATATACAGGCTCTCGCTATGTCGACCATATATGAATTTCTCTGAAGCTTCAAGAA 1416	
QY	384	AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPhe 402	
DB	1417	ACGCTAGGCTCGACGCTGATAAATTTCTGCTGTGATAATGAGCTGTACCCGGATTT 1476	
QY	403	GlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluGluIleLeuAla 422	
DB	1477	TTGTTTACGGCGCATGGAGTAACCA---TTTCAACATATATTGCTGAATTTAAT 1533	
QY	423	LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGly 442	
DB	1534	CACAAACATTCGGTATTATATATATCGGGTGATAAGGATTATATTGTAATTGGCTGGGA 1593	
QY	443	GlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGly 462	
DB	1594	AACCATGCTTGTCCAAATGGATGGATCAATAAACAGTAGGTATCAGACAGAGGATG 1653	
QY	463	TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 480	
DB	1654	TTAAGACCATGGTTCAGTAAAGAAACAGGTGAAGAGTTGGGACAGTCAAGACATATGGC 1713	
QY	481	AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 500	
DB	1714	CCTTTACCTTTTGGAGATATACGATCGCGTCAATGCTGCTCATGATCAACCGGAG 1773	
QY	501	AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 512	
DB	1774	GCAAGTTTGAAATGGTCAACAGTTTCCGGT 1809	

RESULT 15

AX536454 2027 bp DNA linear PAT 22-NOV-2002

LOCUS Sequence 55 from Patent WO02064766.

DEFINITION AX536454

ACCESSION AX536454.1 GI:25262853

VERSION

KEYWORDS Saccharomyces cerevisiae (baker's yeast)

SOURCE Saccharomyces cerevisiae

ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 Contreras, R.H., Eberhardt, I., Luyten, W.H. and Reekmans, R.J.

AUTHORS Bax-responsive genes for drug target identification in yeast and fungi

TITLE Patent: WO 02064766-A 55 22-AUG-2002;

JOURNAL JANSSEN PHARMACEUTICA N.V. (BE)

FEATURES Location/Qualifiers

source 1. 2027

/organism="Saccharomyces cerevisiae"

/mol type="genomic DNA"

/db_xref="taxon:4932"

BASE COUNT 583 a 380 c 427 g 637 t

ORIGIN

Alignment Scores:

Pred. No.: 5,48e-40 Length: 2027

Score: 636.00 Matches: 167

Percent Similarity: 46.29% Conservative: 70

Best Local Similarity: 32.62% Mismatches: 197

Query Match: 21.35% Indels: 78

DB: 6 Gaps: 14

US-09-712-338-2 (1-555) x AX536454 (1-2027)

```

QY 19 LeuProGlySerThrProAlaSerValGlyArgArgGlnLeu-----ProLys 34
Db 627 CTTCCACAGATACACACCAACCTTAATGTCGACCGTTTGAATCAGATGATCCGCTG 686
QY 35 AsnProThrGlyValLeuThrThrAlaAsnAsnValThrIleArgTyrLysGlu 54
Db 687 TTACAACTTTTATTTCTCTGTGTGACACAGATTACAGTTTGAGACTTAGAACAGTAGAT 746
QY 55 ProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal 74
Db 747 CTTCTAACTAGGAATT-----GACACCCGTAAACCAATGTCGGGTATACATG 794
QY 75 AspThrSerProGluSerHisThrPheThrPhePheGluAlaArgHisAsnProGlu 94
Db 795 GACTATAAGGATTCACCAACACTTTTCTTACGGTTTGTGACCTGGTTCTCTGTTTACTGGGTG 854
QY 95 ThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 114
Db 855 AACGACCCCAATATTCTTTGGTTAAATGGTGACCTGGTTCTCTGTTTACTGGGTG 914
QY 115 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis 134
Db 915 CTATTTGAACCTAGGCCCCCTCATCAATTGGCGCCGATATGAAACCAATCCCAATCCCTAT 974
QY 135 SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer 154
Db 975 TCTTGGTAATAACACGCTTCAATGATCTTTAGAACAGCCACTCGGAGTCGGCTTTTTC 1034
QY 155 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer 174
Db 1035 TATGGTGAT----- 1043
QY 175 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsn 194
Db 1044 -----GAAAAAGTCTCTCTCAAAAA 1064
QY 195 LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu 214
Db 1065 TTAGCAGGCAAGATCGGTACATTTCTCTGGAATTTCTTGTGAAAGCTTTTCTCTCATTTA 1124
QY 215 AspSerArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyr 234
Db 1125 -----CGCTCCAACGATTTCCCATGTCAGGCGAATCCCTATGCAGGACATTAT 1172
QY 235 GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer 253
Db 1173 ATCCCTCAAAATGTCACATGAGATCGTTGTCAAGAACCTCGAAGA----- 1217
QY 254 ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGlu 273
Db 1218 -----ACGTTCAATTAACTTCAGTTATGATTGTAATGGTATCAGACCCCT 1265
QY 274 AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys 292
Db 1266 TTGATTCAAGCAGATTATTATGAACCAATGCGATGCGGGAAGGGGGCTATCACCTGTT 1325
QY 293 AlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 312
Db 1326 CTTCTCATCAGAAAGATGTGAGAAAATGAGTAAAGCTGCAGGTCGT----- 1370
QY 313 GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu 332
Db 1371 -----TGTCGTAGTTGAACAAAGTTATGTTATGCTTCTAAATCAAGT 1412
QY 333 -----CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAla 350
Db 1413 TTACCATGCATAGTCGCCACTGCTTACTGTGACTCTGCACCTTTTGAACCCGTACATTAAAC 1472
QY 351 PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro----- 366

```

```

Db 1473 ---ACAGACTCAACGCTCTATGACATTAGAGGGCCCTGTGAGATAAATAGTACTGATGCT 1529
QY 367 -----ThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 383
Db 1530 ATGTTTATACAGGTCTCCGCTATGTCGACAGTATATGAATTTTCCTGAAGTTCAAGAA 1589
QY 384 AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAspValTyrTyrAlaPhe 402
Db 1590 ACGTAGGTCGCGACGTGATATATTTCTGCTGTGATAATGACGTGTTCCACCGGATTT 1649
QY 403 GlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIleLeuAla 422
Db 1650 TTGTTTACGGCGATGAAGTAAACCA---TTTCAACAATATATTTGCTGAATTTAAT 1706
QY 423 LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGly 442
Db 1707 CACAACATTCGGGTATTAAATATATATCGGGTATAGAGTATATATTTGTAATGGCTGGGA 1766
QY 443 GlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGly 462
Db 1767 AACCATGCTTGGTCCAATGAGTTGGAATGGATCAATAAACGTAGGTATCAGAGAAGGATG 1826
QY 463 TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 480
Db 1827 TTAAGACCATGGGTCACTAAAGAAACAGGTGAAGAGTTGGGCAAGTCACAGACTATGGC 1886
QY 481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 500
Db 1887 CCTTTCACCTTTTGTAGATATACGATCGCGTCAATATGCTGCTATGATCAACCGGAG 1946
QY 501 AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 512
Db 1947 GCAAGTTTGGAAATGGTCAACAGTTGGAATTCGGT 1982

```

Search completed: November 21, 2003, 19:36:20

Job time : 4946 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 20:19:24 ; Search time 2609 Seconds
(without alignments)
5170.174 Million cell updates/sec

Title: US-09-712-338-2
Perfect score: 555
Sequence: 1 MRGYEFLSVLPVAAWALP.....HTQSSVPLPATSMSVGMMA 555

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 segs, 12152238056 residues

Word size: 1
Total number of hits satisfying chosen parameters: 45562604

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool_p/US09712338/runat 17112003 170558 11713/app_query.fasta_1.711
-DB=EST -Qfmt=fastap -SUFFIX=oli.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @cgn_1.1.3549 @runat 17112003 170558 11713 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1: em estba:*
2: em esthum:*
3: em estin:*
4: em estmu:*
5: em estov:*
6: em estpl:*
7: em estro:*
8: em htc:*
9: gb est1:*
10: gb est2:*
11: gb htc:*
12: gb est3:*
13: gb est4:*
14: gb est5:*
15: em estfun:*
16: em estom:*
17: em gss_hum:*
18: em gss_inv:*
19: em gss_pln:*
20: em gss_vrt:*
21: em gss_vrt:*
22: em gss_fun:*
23: em gss_mus:*
24: em gss_pro:*
25: em gss_rod:*
26: em gss_phg:*
27: em gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	3.4	588	9	AM672518
2	10	1.8	328	10	BE667001
3	10	1.8	349	10	BE665749
4	10	1.8	398	12	BM481217
5	10	1.8	414	9	AW315989
6	10	1.8	450	9	AI329363
7	10	1.8	479	14	CB438457
8	10	1.8	486	10	BF653688
9	10	1.8	548	10	BF077122
10	10	1.8	656	14	CB419003
11	10	1.8	876	14	CD384438
12	9	1.6	59	28	BH848880
13	9	1.6	185	29	CC409857
14	9	1.6	245	10	BES28387
15	9	1.6	292	9	AA874369
16	9	1.6	311	10	BF472693
17	9	1.6	317	9	AU062698
18	9	1.6	363	29	FR0009619
19	9	1.6	382	9	AA004199
20	9	1.6	383	13	BY137964
21	9	1.6	396	14	C99707
22	9	1.6	407	14	D39358
23	9	1.6	419	13	BQ280268
24	9	1.6	427	12	BM492161
25	9	1.6	445	28	AQ948352
26	9	1.6	447	10	BES22734
27	9	1.6	451	28	AQ647858
28	9	1.6	452	28	BZ178325
29	9	1.6	454	10	BE341016
30	9	1.6	459	14	C73765
31	9	1.6	461	10	BE789144
32	9	1.6	462	12	B3563841
33	9	1.6	469	9	AW648202
34	9	1.6	469	12	B3560934
35	9	1.6	470	10	BG557369
36	9	1.6	476	9	AI331526
37	9	1.6	478	13	BQ584407
38	9	1.6	484	9	AW433393
39	9	1.6	489	29	BZ897787
40	9	1.6	493	14	CB725077
41	9	1.6	521	12	BJ179481
42	9	1.6	521	28	AQ611504
43	9	1.6	532	12	BJ195187
44	9	1.6	533	13	BQ449800
45	9	1.6	555	12	B3563385

ALIGNMENTS

RESULT 1
AW672518
LOCUS
DEFINITION
LGI_360_F03_b1_A002 Light Grown 1 (LGI) Sorghum bicolor CDNA, mRNA
ACCESSION
AW672518
VERSION
AW672518.1
KEYWORDS
EST.
SOURCE
Sorghum bicolor (sorghum)
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAG
Plate: 60 row: C column: 1
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers

FEATURES
source
1..328
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40 embryos."

BASE COUNT 72 a 92 c 94 g 70 t
ORIGIN
Alignment Scores:
Pred. No.: 12.6 Length: 328
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 10 Gaps: 0

US-09-712-338-2 (1-555) x BE667001 (1-328)
QY 6 PheLeuSerValLeuProLeuValAla 15
Db 88 TTCCCTTCGGTTTTCCTCTGCTGGCGGCC 59
RESULT 3
BE665749/c 349 bp mRNA linear EST 25-APR-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE665749
154997 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
BE665749
BE665749.1 GI:10026340
EST.
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE
AUTHORS
1 (bases 1 to 349)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.

TITLE
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
JOURNAL
MEDLINE
PUBMED
COMMENT

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAG
Plate: 68 row: M column: 20

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 588)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L.H.
An EST database from Sorghum: light-grown seedlings
Unpublished
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
POLYA-No.
High quality sequence stop: 580
Location/Qualifiers

FEATURES
source
1..588
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI
The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 145 a 151 c 152 g 140 t
ORIGIN

Alignment Scores:
Pred. No.: 1.97e-08 Length: 588
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.42% Indels: 0
DB: 9 Gaps: 0

US-09-712-338-2 (1-555) x AW672518 (1-588)
QY 99 ThrLeuTrpLeuAsnGlyProGlySerAspSerLeuIleGlyLeuPheGluGlu 117
Db 320 ACTTGTGTTGAATGGCGTCCCGATCAGATCGCTGATGGCTCTTTGAAGAA 376
RESULT 2
BE667001/c 328 bp mRNA linear EST 25-APR-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE667001
151067 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
BE667001
BE667001.1 GI:10027592
EST.
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE
AUTHORS
1 (bases 1 to 328)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.

TITLE
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
JOURNAL
MEDLINE
PUBMED
COMMENT

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

```

Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1. .349
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT      89 a  93 c  98 g  69 t
ORIGIN
Alignment Scores:
Pred. No.:      13.4      Length:      349
Score:          10.00     Matches:      10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:      1.80%      Indels:      0
DB:              10         Gaps:        0

US-09-712-338-2 (1-555) x BE665749 (1-349)
QY      6 PheLeuSerValLeuProLeuValAlaA15
Db      112 TTCTTTCGGTTTGGCTCTGCTGGTGGCGGCC 83

RESULT 4
AW481217
LOCUS      BM481217      398 bp      mRNA      linear      EST 05-FEB-2002
DEFINITION 533117 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  BM481217
VERSION     BM481217.1 GI:18531545
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
ORGANISM    Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 398)
Smith,T.P.L., Grosse,W.M., Preking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
11282978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAG
Plate: 3 row: A column: 18
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1. .398
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT      91 a  132 c  117 g  74 t
ORIGIN
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1. .349
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT      91 a  112 c  113 g  82 t
ORIGIN
Alignment Scores:
Pred. No.:      15.4      Length:      398
Score:          10.00     Matches:      10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:      1.80%      Indels:      0
DB:              12         Gaps:        0

US-09-712-338-2 (1-555) x BM481217 (1-398)
QY      6 PheLeuSerValLeuProLeuValAlaA15
Db      257 TTCTTTCGGTTTGGCTCTGCTGGTGGCGGCC 286

RESULT 5
AW315989/c
LOCUS      AW315989      414 bp      mRNA      linear      EST 25-APR-2001
DEFINITION 13811 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  AW315989
VERSION     AW315989.1 GI:6745236
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
ORGANISM    Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 414)
Smith,T.P.L., Grosse,W.M., Preking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
11282978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAG
Plate: 10 row: F column: 9
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1. .414
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT      91 a  132 c  117 g  74 t
ORIGIN

```

RESULT 7
CB438457 479 bp mRNA linear EST 25-MAR-2003
LOCUS
DEFINITION 687060 MARC 6BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION CB438457
VERSION CB438457.1 GI:29222489
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 479)
AUTHORS Smith, R.P.L., Roberts, A.J., Echtenkamp, S.E., Chitko-McKown, C.G.,
Wray, J.E. and Keele, J.W.
A second set of bovine ESTs from pooled-tissue normalized libraries
Unpublished
Contact: Smith TPL
USDA, ARS US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: FQ18043 row: P column: 7
Seq primer: GTATACGACTCAGCATAGGG.
Location/Qualifiers
1..479
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 6BOV"
/note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
library made with RNA pooled from multiple tissues
including liver, lung, hypothalamus, pituitary, and
placenta/endometrium."
BASE COUNT 124 a 143 c 134 g 78 t
ORIGIN
Alignment Scores:
Pred. No.: 18.8 Length: 479
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 14 Gaps: 0
US-09-712-338-2 (1-555) x CB438457 (1-479)
Qy 6 PheLeuSerValLeuProLeuValAlaAla 15
Db 118 TTCCTTCGCTTTTGCTCTGTCGGCGGCC 147
RESULT 8
BF653688 486 bp mRNA linear EST 25-APR-2001
LOCUS
DEFINITION 277602 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF653688
VERSION BF653688.1 GI:11918820
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 486)
AUTHORS Smith, R.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.M., Rohrer, G.A., Chitko-McKown, C.G.,

Alignment Scores:
Pred. No.: 16.1 Length: 414
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 9 Gaps: 0
US-09-712-338-2 (1-555) x AW315989 (1-414)
Qy 6 PheLeuSerValLeuProLeuValAlaAla 15
Db 414 TTCCTTCGCTTTTGCTCTGTCGGCGGCC 385
RESULT 6
AI329363 450 bp mRNA linear EST 28-DEC-1998
LOCUS
DEFINITION b3g1one.f1 Neurospora crassa evening cDNA library Neurospora crassa
cDNA clone b3g1one 5', mRNA sequence.
ACCESSION AI329363
VERSION AI329363.1 GI:4065922
KEYWORDS EST.
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 (bases 1 to 450)
AUTHORS Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
TITLE Two Neurospora crassa EST Databases
JOURNAL Unpublished
COMMENT Other ESTs: b3g1one.r1
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 353.
Location/Qualifiers
1..450
/organism="Neurospora crassa"
/mol_type="mRNA"
/strain="Strain 30-7 (bd; A)"
/db_xref="taxon:5141"
/clone="b3g1one"
/tissue_type="tissue harvested following 22hr growth in
dark"
/clone_lib="Neurospora crassa evening cDNA library"
/note="Vector: pBluescript SK-; Site 1: XbaI; Site 2:
EcoRI; See: Bell-Perderson, D., et al. PNAS 93:13095, 1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3'
end of cDNA cloned into EcoRI site of pBluescript"
BASE COUNT 93 a 147 c 112 g 98 t
ORIGIN
Alignment Scores:
Pred. No.: 17.6 Length: 450
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 9 Gaps: 0
US-09-712-338-2 (1-555) x AI329363 (1-450)
Qy 227 ThrGluSerTyrGlyGlyHisTyrGlyPro 236
Db 401 ACGGAAGTTATGGGCGGCATACGGGCT 430

Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
21182978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -muncore 18 and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 66 row: G column: 21
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
source
1..486
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH103"
/clone_lib="MARC 3BOV"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."
BASE COUNT 100 a 136 c 155 g 95 t
ORIGIN

Alignment Scores:
Pred. No.: 19.1 Length: 486
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 10 Gaps: 0

US-09-712-338-2 (1-555) x BF653688 (1-486)

QY 6 PheLeuSerValLeuProLeuValIalaLa 15
Db 368 TTCCCTTCGGTTTTCCTCTGGTGGCGCC 397

RESULT 9
BF077122 548 bp mRNA linear EST 25-APR-2001
LOCUS
DEFINITION 228926 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF077122
VERSION BF077122.1 GI:10870952
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 548)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,B., Wray,J.E., White,J., Cho,J., Fahrnkug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
21182978

DB:	14	Gaps:	0
US-09-712-338-2 (1-555) x CD384438 (1-876)			
Qy	228	GlusertyrGlyHisTyrGlyProAla 237	
Db	255	GAATCCTACGCTGGACACTATGTCGCGCC 284	
RESULT 12			
BH848880			
LOCUS			
DEFINITION	BH848880	59 bp DNA	linear GSS 13-JUN-2002
	SALK_068974.20.80.x Arabidopsis thaliana TDNA insertion lines		
	Arabidopsis thaliana genomic clone SALK_068974.20.80.x, genomic		
	survey sequence.		
ACCESSION	BH848880		
VERSION	BH848880.1	GI:21419751	
KEYWORDS	GSS.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids		
	; eurosids II; Brassicales; Brassicaceae; Arabidopsi.		
REFERENCE	1 (bases 1 to 59)		
AUTHORS	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab		
	, C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.		
	, Zimmermann,J. and Ecker,J.R.		
TITLE	A Sequence-indexed Library of Insertion Mutations in the		
JOURNAL	Arabidopsis Genome		
COMMENT	Unpublished		
	Contact: Joseph R. Ecker		
	Salk Institute Genomic Analysis Laboratory (SIGAL)		
	The Salk Institute for Biological Studies		
	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA		
	Tel: 858 453 4100 x1752		
	Fax: 858 558 6379		
	Email: ecker@salk.edu		
	This is single pass sequence recovered from the left border of		
	TDNA.		
FEATURES	Class: TDNA tagged.		
source	Location/Qualifiers		
	1..59		
	/organism="Arabidopsis thaliana"		
	/mol_type="genomic DNA"		
	/strain="Columbia 0"		
	/db_xref="taxon:3702"		
	/clone="SALK_068974.20.80.x"		
	/clone_lib="Arabidopsis thaliana TDNA insertion lines"		
	/note="PCR was performed on Arabidopsis thaliana lines		
	each of which contains one or more TDNA insertion		
	elements. The resultant fragment for each line was		
	directly sequenced to determine the genomic sequence at		
	the site of insertion. Details of the protocols used can		
	be found at http://signal.salk.edu/tdna_protocols.html "		
BASE COUNT	11 a 24 c 10 g 14 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	20.5	Length:	59
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.62%	Indels:	0
DB:	28	Gaps:	0
US-09-712-338-2 (1-555) x BH848880 (1-59)			
Qy	543	LeuProThrAlaThrSerMetSerSer 551	
Db	10	TTCGCCACGGCCACGTCATGTCCTCC 36	
RESULT 13			
CC409857/c			
LOCUS	CC409857	185 bp	DNA linear GSS 19-MAY-2003


```

DEFINITION PUHCZ89TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa423010,
genomic survey sequence.
ACCESSION CC409857
VERSION CC409857.1 GI:30889947
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
REFERENCE 1 (bases 1 to 185)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
Maize Genomics Consortium
TITLE A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
JOURNAL Unpublished
COMMENT Other GSSs: PUHCZ89T8
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
source
1..185
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa423010"
/clone_lib="ZM_0.6_1.0_KB"
/notes="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"
BASE COUNT 59 a 49 c 55 g 22 t
ORIGIN
Alignment Scores:
Pred. No.: 69.5 Length: 245
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 29 Gaps: 0
US-09-712-338-2 (1-555) x CC409857 (1-185)
QY 8 ServValLeuProLeuValAlaLaSer 16
Db 184 TCTGTGCTGCCGCTGGTCGCTGGTCA 158
RESULT 14
LOCUS BE528387
DEFINITION M79C20STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION BE528387
VERSION BE528387.1 GI:9786365
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
1 (bases 1 to 245)
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
JOURNAL 20567808
MEDLINE 11115876
PUBMED
COMMENT Contact: Benning, C

```

```

Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
USA, FAX: 6142920603 TEL: 6142929371.
FEATURES
source
1..245
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="600037474R1"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/clone_lib="Arabidopsis developing seed"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
BASE COUNT 54 a 53 c 66 g 72 t
ORIGIN
Alignment Scores:
Pred. No.: 93.7 Length: 245
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 10 Gaps: 0
US-09-712-338-2 (1-555) x BE528387 (1-245)
QY 99 ThrLeuTrpLeuAsnGlyGlyProGly 107
Db 121 ACCCTTTGGCTCAATGGAGGTCCAGGT 147
RESULT 15
LOCUS AA874369/c
DEFINITION vx07g10.r1 Soares thymus 2MbMT Mus musculus cDNA clone
IMAGE:1263810 5', similar to TR:O35952 O35952 RSP29. ;, mRNA
sequence.
ACCESSION AA874369
VERSION AA874369.1 GI:2979058
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 232)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:666362
Possible reversed clone: similarity on wrong strand

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 19:36:24 ; Search time 4885 Seconds
(without alignments)
4647.870 Million cell updates/sec

Title: US-09-712-338-2

Perfect score: 555

Sequence: 1 MRGYEFLSVLPVAASWALP.....HTQSSVPLPTATSMSSVGMMA 555

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5773148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09712338/runat_17112003_170558_11701/app_query.fasta_1.711
-DB=GenEmbl -QFMT=fastap -SUFFIX=oli.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CGN 1 1 4958 @runat_17112003_170558_11701 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rod:*
36: em.htg.mam:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	554	99.8	1662	6	AR129928 Sequence
2	191	34.4	2245	8	AF394242 Aspergill
3	24	4.3	1656	6	AX534871 Sequence
4	24	4.3	3150	6	AX534814 Sequence
5	10	1.8	1581	6	AX534824 Sequence
C 6	10	1.8	1883	8	IPBSPBA
7	10	1.8	2940	6	AX534767
C 8	10	1.8	4772	8	IPBAMYB
C 9	10	1.8	30855	2	AC025720
10	10	1.8	46000	3	AC084156
C 11	9	1.6	144	9	S63789
12	9	1.6	299	6	AR250918
C 13	9	1.6	467	1	KSP302770
C 14	9	1.6	571	1	PPU296313
15	9	1.6	1422	8	AY143899
16	9	1.6	1441	8	BT004386
17	9	1.6	1527	6	AX594884
18	9	1.6	1654	8	BT004306
19	9	1.6	1665	6	AX534866
20	9	1.6	1690	8	AF361604
21	9	1.6	1704	8	BT002948
22	9	1.6	1872	6	AX534872
23	9	1.6	2016	8	SCYBR139W
24	9	1.6	2027	6	AX536454
C 25	9	1.6	2506	8	BT002365
C 26	9	1.6	2722	8	AY091014
C 27	9	1.6	2910	8	BT005734
28	9	1.6	3080	6	AX534809
C 29	9	1.6	3213	9	HSEXMONTR
30	9	1.6	3221	6	AX534815
31	9	1.6	3298	8	ABGSCP2
C 32	9	1.6	3346	1	PSU81032
33	9	1.6	3859	4	AF081464
34	9	1.6	7827	1	KAR751A
C 35	9	1.6	9566	1	RP1W17A
C 36	9	1.6	11123	1	AF251288
C 37	9	1.6	12373	1	PSF17897
C 38	9	1.6	19295	1	AF498414
39	9	1.6	21200	8	AP002044
C 40	9	1.6	22913	1	AF498407
41	9	1.6	29686	8	SC29711
C 42	9	1.6	37288	3	CEC0505
43	9	1.6	50277	2	AC138524_5
C 44	9	1.6	53425	1	EA067194
45	9	1.6	55216	6	AR306445

ALIGNMENTS

RESULT 1

AR129928 1662 bp DNA linear PAT 16-MAY-2001
LOCUS
DEFINITION Sequence 1 from patent US 6187578.
ACCESSION AR129928
VERSION AR129928.1 Gi:14117825
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE UNCLASSIFIED.
AUTHORS 1 (bases 1 to 1662)
Blinkovsky, A., Berka, R., Rey, M., Golightly, E., Klotz, A.,
Mathisen, T., Erik., Dambmann, C. and Brown, K.M.
TITLE Carboxypeptidases and nucleic acids encoding the same
JOURNAL Patent: US 6187578-A 1 13-FEB-2001;
FEATURES Location/Qualifiers
1..1662
/organism="unknown"
BASE COUNT 396 a 468 c 398 g 400 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1662
Score: 554.00 Matches: 554
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.82% Indels: 0
D8: 6 Gaps: 0
US-09-712-338-2 (1-555) x AR129928 (1-1662)
QY 1 MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuPro 20
Db 1 ATGCGTGGTACGAATTTCTCTCAGTGTCTACCCCTTGTTGCAGCCAGTTGGGCCCTTCCA 60
QY 21 GlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGlyValIlys 40
Db 61 GGAAGTACACCGCGCTCGGTGTAGACAGACTACCAAGAACCCACCGGGGTCAAG 120
QY 41 ThrLeuThrThralaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal 60
Db 121 ACTCTTACACCCGCAACATGTCACATCCGTTACAGAACCCGGGCGAGGGCGTC 180
QY 61 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer 80
Db 181 TGGAGACTACCCCGGGTGTCAATCTCTCTGTGATATGTCGACACCTCTCCCGAGTCC 240
QY 81 HisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
Db 241 CATACCTTCTCTGGTTCTTCGAAGCCAGACATACCCAGAACTGCACCTATCACATTG 300
QY 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 120
Db 301 TGGTTGAATGGTGGCCCTGAAGAGGATCTTTGATCGTCTCTTCGAAGAGTTGGGCCCT 360
QY 121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 140
Db 361 TGCATGTCAATTCGATTTGATGACTACATCAACCTCCTCCTCGTGGAACTGAGGCTCC 420
QY 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
Db 421 AATTACTATTCTCTCCAGCCACTGGGAGTCTTTGATCGTCTCTTCATATGATGATACGGTAT 480
QY 161 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 180
Db 481 GGGTCCATTACCTCTGAACTGGGGTCTGCGAAATTCGAGCTTTGCAGGAGTTTCAGGGC 540
QY 181 ArgTyrProThrIleAspAlaThrIleuIleAspThrThrAsnLeuAlaAlaGluAlaAla 200
Db 541 CGGTACCCCAACCATTTGATGACCTCTGATCATCTACCAATCTTGCAGAGAGGGCGCT 600
QY 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
Db 601 TGGAGATCTCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTGCAGTCT 660

QY 221 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn 240
Db 661 AAGGACTTCAGTCTATGACGAGAGCTATGAGGGCAGTATGTCCTGCAATTCCTCAAT 720
QY 241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
Db 721 CATTTTACGAGCAGAAATGAGAGAAATGCCAACGGTAGTGTAAATGGTGTTCAGCTTAAT 780
QY 261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 280
Db 781 TTCACCTCTCTGGGAATATTAAACGGCATCATCGACGAGCGATCCAGGCCCTTACTAC 840
QY 281 ProGluPheAlaValAsnAsnThrTyrGlyIleIysAlaValAsnGluThrValTyrAsn 300
Db 841 CCTGAATTCGCTGTGAACATACCTACGTTACAGGTGTCAAGGACGCTCAGACACCTCAGAAC 900
QY 301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
Db 901 TACATGAAGTTTGGCCAAACCAATGCCAATGGTTCGCCAGGATTTGATTTCCACCTGCAAA 960
QY 321 GlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys 340
Db 961 CAGACAAACCGCACCGCATTAGCTGACTACGCCCTCTGCGCGAAGCCACCAACATGTGC 1020
QY 341 ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 360
Db 1021 AGGGACAATGTTGAGGGGCCATACACGCTTTGCTGTGCTGTGGTGTGTATGATATTCGG 1080
QY 361 HisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSer 380
Db 1081 CATCCATATGATGACCCGACTCCGCCAAGTTATTACACAAATTTCTGCAAGGACTCT 1140
QY 381 ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 400
Db 1141 GTCATGGACGCTATCGGCGTCAACATCAACTACCCAGTCCCAATATGACGCTACTAC 1200
QY 401 AlaPheGlnGlnThrGlyAspPheValTyrProAsnPheIleGluAspLeuGluGluIle 420
Db 1201 GCTTTCCAGCAACAGCGGACTTTGCTGGGCCCACTTCATCGAAGACCTCGAGGAGATC 1260
QY 421 LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp 440
Db 1261 CTTGCTCTCCCGTGGTGTCTCCCTCATCTATGGCGACGCGGATTCATCTCAACTGG 1320
QY 441 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 460
Db 1321 TTCGGCGGTGAGCCGCTTTCCCTCGCTGCGAACTACTCCCAAGCCGCCAGTTCGGAAGC 1380
QY 461 AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 480
Db 1381 GCAGGGTACACGCCCTCGAAGTCAACGGCGTCAACGGCGTCAAGTCCCATCTACCAACCCATC 1440
QY 481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 500
Db 1441 AATTTCTCTCTCACTCGCTCTATGAGCAGCGCATGAAGTCCCATCTACTACCAACCCATC 1500
QY 501 AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys 520
Db 1501 GCCTCCCTGCAATTTGTTAAACCGGACTATCTTCGTTGGGATATCGCAGAGGGCCAGAG 1560
QY 521 LysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 540
Db 1561 AAGATCTGGGCCCACTACAGACGAATGGACGGCTACAGCTACGATACACATACACAGTCTCGTCC 1620
QY 541 ValProLeuProThrAlaThrSerMetSerSerValGlyMet 554
Db 1621 GTGCGCTGCTACGGCTACCGCATGTCAGCATGTCAGTCTTGGTATG 1662

RESULT 2
AF394242 2245 bp DNA linear PLN 24-JUL-2001
LOCUS AF394242
DEFINITION Aspergillus oryzae strain TK3 carboxypeptidase S1 (cpl) gene,

```
complete cds.
ACCESSION AF394242
VERSION AF394242.1 GI:15004615
KEYWORDS
SOURCE
ORGANISM
Aspergillus oryzae
Aspergillus oryzae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE
1 (bases 1 to 2245)
AUTHORS van den Broek, P.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,
Lausanne CH-1000, Switzerland
FEATURES
Location/Qualifiers
1..2245
/organism="Aspergillus oryzae"
/mol_type="genomic DNA"
/strain="TK3"
/db_xref="taxon:5062"
<1..>2245
/gene="cpi"
join(<1..349,409..513,576..692,743..866,941..956,
1008..1124,1187..1237,1296..1449,1522..1580,1670..>2245)
/gene="cpi"
/product="serine carboxypeptidase 1"
join(1..349,409..513,576..692,743..866,941..956,
1008..1124,1187..1237,1296..1449,1522..1580,1670..>2245)
/gene="cpi"
/EC number="3.4.16.6"
/codon_start=1
/product="carboxypeptidase S1"
/protein_id="AAK7166.1"
/db_xref="GI:15004616"
/translation="MRGYEFLSVLPLVAASWALPGSTPASVGRQLPKNPTGVKTLIT
ANNVTRYKEPGAEGVCEITPGVKSQSVYDVSPESTFFWFPEARHNPETAPITLWL
NGPGQSDLLIGLFEELGPCHVSTFDYINPHSWNEVSNLLFSLQPLGVGFSYNDTVD
GSPINPTVWNSFAGVQGRYPTDITLDTNLAEEAAWEILQFLSGLSLDSRV
QSKDPSLWNTSGHYGPAFFNFHYEQNERIANGSVNGVQLNFNSLIGLIDEAIO
APYEPFANVTYGVKAVNETVYMKFQNPQNGODLITCKOTNRALADYALCA
EATNCRDNVEGYPYAFAGRVVDIRHPVDPTPPSYNKFELAKSVMDAIGNVNT
QSNNDVYAFQOTGDFWNFLEDEILALPVRSLYICNWFEGQAVSLAA
NYSQAQFRSAGTYELKVNVEYGETRENFSTRVYBAGHEVPYQPIASLQLFNR
TIFGMDIAGQKKIWPXYKNTGTATATHTQSSVPLPTATSMNSVGMA"
sig_peptide
1..54
/gene="cpi"
BASE COUNT 548 a 603 c 524 g 570 t
ORIGIN
Alignment Scores:
Pred. No.: 1.14e-195 Length: 2245
Score: 191.00 Matches: 191
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 34.41% Indels: 0
DB: 8 Gaps: 0
US-09-712-338-2 (1-555) x AF394242 (1-2245)
QY 365 AspProThrProSerTyrThrAsnLysPheLeuAlaLysAspSerValMetAspAla 384
Db 1670 GACCCGACTCCGCCAAGTTATTACACAAATTTGGCCAGGACTCTGTCTATGACGCT 1729
QY 385 IleGlyValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyrAlaPheGlnGln 404
Db 1730 ATCGCGGTCAACATCACTACACCCAGTCCATATGAGCTCTACTACGCTTTCAGCAA 1789
QY 405 ThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIleLeuAlaLeuPro 424
Db 1790 ACAGCGCATTTGTCTGGCCCACTTCATCGAAGACCTCGAGAGATCTTCTCTCCC 1849
QY 425 ValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGln 444
Db 1850 GTGCGTGTCTCCCTCATCTATGGCAGCCGCGATTACATCTGCAACTGTGTCGCGGTG 1909
```

```
QY 445 AlavalSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrThr 464
Db 1910 GCCGTTTCCCTCGCTCGAACTACTCCAAAGCCGCCAGTTCCGAAGCGAGGTACACG 1969
QY 465 ProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPhe 484
Db 1970 CCCTGAAAGTCAACGCGTCGAGTATGGGAACTCGCGAGTATGTAATTTCTCGTTC 2029
QY 485 ThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGln 504
Db 2030 ACTCGCGTCTATGAGCAGGCCCATGAGTCCCATCTACAGGCCCATCGCTCCCTGCAA 2089
QY 505 LeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGlyGlnLysIleTyrPro 524
Db 2090 TTGTTTAAACCGGACTATCTTCGGTTGGGATATCGCAGAGGCCAGAGAAGATCTGSCC 2149
QY 525 SerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValProLeuPro 544
Db 2150 AGCTACAGACGAATGGAACGGCTACAGCTACGCATACACAGTCGTCGCGTGCCT 2209
QY 545 ThrAlaThrSerMetSerSerValGlyMetAla 555
Db 2210 ACGGCTACCGCATGTCAGTGTGTGATGGCT 2242
RESULT 3
LOCUS AX534871
DEFINITION Sequence 109 from Patent WO02068623.
ACCESSION AX534871
VERSION AX534871.1 GI:25261462
KEYWORDS
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1
AUTHORS Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A.,
Kimpel, B., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,
Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,
Kemmer, W., Tan, P., Stiebler, J. and Albang, R.
TITLE Novel genes encoding novel proteolytic enzymes
JOURNAL Patent: WO 02068623-A 109 06-SEP-2002;
DSM N.V. (NL)
FEATURES
Location/Qualifiers
1..1656
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT 411 a 426 c 409 g 410 t
ORIGIN
Alignment Scores:
Pred. No.: 2.2e-15 Length: 1656
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.32% Indels: 0
DB: 6 Gaps: 0
US-09-712-338-2 (1-555) x AX534871 (1-1656)
QY 99 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGlnLeu 118
Db 295 ACTCTGTGCTGAATGGTGGCCCTGGAAGCGATTCTTGTATGGCTTTTGAAGAGTTG 354
QY 119 GlyProCysHis 122
Db 355 GGTCCGTGTCAC 366
RESULT 4
LOCUS AX534814
DEFINITION Sequence 3150 bp DNA linear PAT 22-NOV-2002
```

Sequence 52 from Patent WO02068623.
ACCESSION AX534814
VERSION AX534814.1 GI:25261343
KEYWORDS Aspergillus niger
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
AUTHORS Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A., Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., Gustedt Von,W., Heinrich,O., Maier,D., Spreafico,F., Folkers,U., Hopper,S., Kemmer,W., Tan,P., Stiebler,J. and Albang,R.
TITLE Novel genes encoding novel proteolytic enzymes
JOURNAL Patent: WO 02068623-A 52 06-SEP-2002;
DSM N.V. (NL)
FEATURES Location/Qualifiers
source 1..3150
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT 801 a 778 c 738 g 833 t
ORIGIN
Alignment Scores:
Pred. No.: 3,99e-15 Length: 3150
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.32% Indels: 0
DB: 6 Gaps: 0
US-09-712-338-2 (1-555) x AX534814 (1-3150)
QY 398 ValTyrTyAlaPheGlnGlnThrGlyAspPheValTrpProAsnPhelLeuAspLeu 417
Db 2295 GTATATTATGCAATCCAGACCGCGACTTGTATGCGCAATTCATTGAGGACCTC 2354
QY 418 GluGluLeuLeu 421
Db 2355 GAAGATCTC 2366
RESULT 5
LOCUS AX534824 1581 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 62 from Patent WO02068623.
ACCESSION AX534824
VERSION AX534824.1 GI:25261366
KEYWORDS Aspergillus niger
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
AUTHORS Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A., Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., Gustedt Von,W., Heinrich,O., Maier,D., Spreafico,F., Folkers,U., Hopper,S., Kemmer,W., Tan,P., Stiebler,J. and Albang,R.
TITLE Novel genes encoding novel proteolytic enzymes
JOURNAL Patent: WO 02068623-A 62 06-SEP-2002;
DSM N.V. (NL)
FEATURES Location/Qualifiers
source 1..1581
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT 343 a 473 c 423 g 342 t
ORIGIN
Alignment Scores:
Pred. No.: 2.81 Length: 1581
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 6 Gaps: 0
US-09-712-338-2 (1-555) x AX534824 (1-1581)
QY 227 ThrGluSerTyrGlyGlyHisTyrGlyPro 236
Db 541 ACGGAGAGTTATGGCGCCACTACGGGCC 570
RESULT 6
LOCUS IPBSPBA 1883 bp mRNA linear PLN 02-FEB-1999
DEFINITION Ipomoea batatas mRNA for beta-amylase, complete cds.
ACCESSION D01022
VERSION D01022.1 GI:217939
KEYWORDS beta-amylase.
SOURCE Ipomoea batatas (sweet potato)
ORGANISM Ipomoea batatas
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; lamids; Solanales; Convolvulaceae; Ipomoea.
REFERENCE 1 (bases 1 to 1883)
AUTHORS Yoshida,N. and Nakamura,K.
TITLE Molecular cloning and expression in Escherichia coli of cDNA encoding the subunit of sweet potato beta-amylase
JOURNAL J. Biochem. 110 (2), 196-201 (1991)
MEDLINE 92105047
PUBMED 1837016
COMMENT Data kindly submitted in computer readable form by: Kenzo Nakamura
Laboratory of Biochemistry
School of Agriculture, Nagoya University
Chikusa-ku, Nagoya
Aichi 464-01
Japan
Phone: 052-781-5111 x6314
Fax: 052-782-3162
Location/Qualifiers
source 1..1883
/organism="Ipomoea batatas"
/mol_type="mRNA"
/strain="Kokei No.14"
/db_xref="taxon:4120"
/clone="pSPbetaA6"
82..1581
/codon_start=1
/product="beta-amylase"
/protein_id="BAA00828.1"
/db_xref="GI:217940"
/translation="MAPIGVMEIGNVSLVYVMLPLGVVNAQNVFPDKKVRDELQKV
KAGGCGVNDVWVGIIKANGPKQYDWSAYRELFLQVKKCGI.KIQAIMSPHOCGNVG
DAVFIPQWILQIGDKNFIYTNAGNRNQEILSLGVNDKQLFGORTALEMYRDFM
ESFRDNMADFLKAGDVIDEVGCGAGELRYSYPTQGWVFGIGFQCYDKYMWAD
WKEAVKQAGNADWMEKPGKAGTYNDTDFKTFPRNGTLQDGYGVFLTWYSNKLIIH
GQOVLTEANKVFLGVNIAAKVGIHWNVHVAHELTAAGVFNAGRGVPIARM
LARHNAITNFTCLEWDESPQAEAKAPQELVQOVLRSQVKEVIDVAGENALPRYDAT
ANQMLKLPNGVNLNGPKLKMSSLTYLRISDDLQTDNFELEFKKFKVKKMEADLDP
SPNATSPAVLERSNATIDELMEATKGRSPFPWVIDVDMFVDCSNPFD"
BASE COUNT 519 a 427 c 486 g 451 t
ORIGIN
Alignment Scores:
Pred. No.: 3.3 Length: 1883
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 8 Gaps: 0
US-09-712-338-2 (1-555) x IPBSPBA (1-1883)
QY 319 CysLysGlnThrAsnArgThrAlaLeuAla 328
|||||

Db 304 TGTAAGCAGACCAATCGTACTGCTTTGGCC 275

RESULT 7
 LOCUS AX534767 2940 bp DNA linear PAT 22-NOV-2002
 DEFINITION Sequence 5 from Patent WO02068623.
 ACCESSION AX534767
 VERSION AX534767.1 GI:25261241
 KEYWORDS
 SOURCE Aspergillus niger
 ORGANISM Aspergillus niger
 Eukaryota; Fungi; Ascomycota; Peizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE
 1 Edens, L., dijk Van A.A., Krubasik, P., Albermann, K., Stock, A.,
 Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,
 Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,
 Kemmer, W., Tan, P., Stiebner, J. and Albang, R.
 Novel genes encoding novel proteolytic enzymes
 Patent: WO 02068623-A 5 06-SEP-2002;
 DSM N.V. (NL)

FEATURES
 source Location/Qualifiers
 1..2940
 /organism="Aspergillus niger"
 /mol_type="genomic DNA"
 /db_xref="taxon:5061" 699 t

BASE COUNT 691 a 818 c 732 g 699 t

ORIGIN

Alignment Scores:
 Pred. No.: 4.99 Length: 2940
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.80% Indels: 0
 DB: 6 Gaps: 0

US-09-712-338-2 (1-555) x AX534767 (1-2940)

QY 227 ThrGluSerTyrGlyGlyHisTyrGlyPro 236
 |||||
 Db 1163 ACGAGAGATTATGGCGCCACTAGCGGCC 1192

RESULT 8
 IPBAMYB
 LOCUS 4772 bp DNA linear PLN 14-APR-2000
 DEFINITION Sweet potato beta-amylase gene, complete cds.
 ACCESSION D12882
 VERSION D12882.1 GI:217935
 KEYWORDS beta-amylase.
 SOURCE Ipomoea batatas (sweet potato)
 ORGANISM Ipomoea batatas
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; lamiids; Solanales; Convolvulaceae; Ipomoea.
 1 (bases 1 to 4772)
 Yoshida, N., Hayashi, K. and Nakamura, K.
 A nuclear gene encoding beta-amylase of sweet potato
 Gene 120 (2), 255-259 (1992)
 MEDLINE 93013042
 PUBMED 1383095

REFERENCE
 2 (bases 1 to 4772)
 Nakamura, K.
 Direct Submission
 Submitted (12-AUG-1992) Kenzo Nakamura, Nagoya University, School
 of Agricultural, Fermentation Technology; Chikusa, Nagoya, Aichi
 464-01, Japan (Tel:052-781-5111(ex.6313), Fax:052-782-9162)
 Submitted (12-AUG-1992) to DDBJ by:
 Kenzo Nakamura
 Lab. of Biochem
 School of Agriculture
 Nagoya University
 Furo-cho, Chikusa-ku

COMMENT

Nagoya 464-01
 Japan
 Phone: 052-781-5111 x6313
 Fax: 052-782-9162.
 Location/Qualifiers
 1..4772
 /organism="Ipomoea batatas"
 /mol_type="genomic DNA"
 /strain="Kokei No.14"
 /db_xref="taxon:4120"
 1236..1242
 1268..1414
 /number=1
 join(1352..1414,1609..2007,2162..2326,2415..2681,
 2763..2972,3727..3966,4087..4242)
 /gene="beta-Amy"
 join(1352..1414,1609..2007,2162..2326,2415..2681,
 2763..2972,3727..3966,4087..4242)
 /gene="beta-Amy"
 /EC_number="3.2.1.2"
 /codon_start=1
 /product="beta-amylase"
 /protein_id="BAA02286.1"
 /db_xref="GI:217936"
 /translation="MAPIPGVMPFGNYSVLYVMLPLGVNADNVFPDKKVEDELKQV
 KAGCGDGVVDVWVWGIEAKPKQYDWSAYRELFQVKGKGLKIQAIMSFQCGNVG
 DAVFPIPIQILQIDKNPDIFTYNNRAGNRNQEYLSLGVNQRLLFQRTALEMYRDM
 ESFRNMADFLKAGDIVDIEVGGAGELRYPSPYPTQGVFPFGIFGFCYDKYVAD
 WKEAVKQAGNADWEMPKGTGYNDTPDKTEFRNGTKYKTMGKFFLTWYKNLLIHH
 GDOVLBEANKVFLGVNIAAKVSGIHWVNHVSHAAELTAGYNNAGRGYRFAEM
 LARHATLNTCLEMRDSEQPAEAKAPQELVOQLSSGWKEYIDVAGENALPYDAT
 AYNQIILNVRPNVNGNPPKIKMSGLTYRLSDLLQTDNFELFKFKVKKMHADLDP
 SPNAISPAVLRSNSAITIDELMEATKGSRPFPFYDVTMPVDGSPNPFD"
 1415..1608
 /gene="beta-Amy"
 /number=1
 1609..2007
 /gene="beta-Amy"
 /number=2
 2008..2161
 /gene="beta-Amy"
 /number=2
 2162..2326
 /gene="beta-Amy"
 /number=3
 2327..2414
 /gene="beta-Amy"
 /number=3
 2415..2681
 /gene="beta-Amy"
 /number=4
 2682..2762
 /gene="beta-Amy"
 /number=4
 2763..2972
 /gene="beta-Amy"
 /number=5
 2973..3726
 /gene="beta-Amy"
 /number=5
 3727..3966
 /gene="beta-Amy"
 /number=6
 3967..4086
 /gene="beta-Amy"
 /number=6
 4087..4546
 /number=7
 BASE COUNT 1480 a 861 c 978 g 1453 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.81 Length: 4772

Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 8 Gaps: 0

US-09-712-338-2 (1-555) x IPBAMYB (1-4772)

Qy 319 CysLysGlnThrAsnArgThrAlaLeuAla 328
Db 1768 TGTAAAGCAGCAATCGTACTGCTTGGCC 1739

RESULT 9
AC025720/c
LOCUS AC025720 30855 bp DNA linear HTG 12-MAR-2000
DEFINITION Caenorhabditis elegans clone Y46E12B, *** SEQUENCING IN PROGRESS
ACCESSION AC025720
VERSION AC025720.1 GI:7230344
KEYWORDS HTG; HTGS PHASE1.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 30855)
AUTHORS Waterston,R.H.
TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 30855)
TITLE Direct Submission
AUTHORS Waterston,R.H.
JOURNAL Submitted (12-MAR-2000) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20853: contig of 20853 bp in length
* 20854 20953: gap of unknown length
* 20954 23864: contig of 2911 bp in length
* 23865 23964: gap of unknown length
* 23965 30855: contig of 6891 bp in length.

FEATURES
source
1. 30855
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/db_xref="taxon:6239"
/clone="Y46E12B"

BASE COUNT 10094 a 5536 c 5703 g 9284 t 238 others
ORIGIN

Alignment Scores:
Pred. No.: 44.1 Length: 30855
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 2 Gaps: 0

US-09-712-338-2 (1-555) x AC025720 (1-30855)
Qy 501 AlaSerLeuGlnLeuPheAsnArgThrIle 510
Db 1561 GCTAGTCTGCAACTTTTAAATAGAACATC 1532

RESULT 10
AC084156
LOCUS AC084156 46000 bp DNA linear INV 22-NOV-2002

DEFINITION Caenorhabditis elegans cosmid Y46E12BL, complete sequence.
AC084156
VERSION AC084156.2 GI:21913070
KEYWORDS HTG.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 46000)
AUTHORS Waterston,R.
TITLE Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
9851916
2 (bases 1 to 46000)
Madsen,C. and Du,H.
The sequence of C. elegans cosmid Y46E12BL
Unpublished (2001)
3 (bases 1 to 46000)
Waterston,R.
Direct Submission
Submitted (13-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 46000)
Waterston,R.
Direct Submission
Submitted (28-JUN-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
5 (bases 1 to 46000)
Waterston,R.
Direct Submission
Submitted (25-APR-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
6 (bases 1 to 46000)
Waterston,R.
Direct Submission
Submitted (20-JUL-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
7 (bases 1 to 46000)
Waterston,R.
Direct Submission
Submitted (22-NOV-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
On Jul 20, 2002 this sequence version replaced gi:10800366.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
email: rw@nematode.wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one m13 subclone.
For a graphical representation of this cosmid sequence and its
analysis see:
http://www.wormbase.org/db/seq/sequence?name=Y46E12BL;class=Sequenc
e

NEIGHBORING COSMID INFORMATION

The 5' cosmid is Y46E12BL, 200 bp overlap.

NOTES:

Coding sequences below are the result of integration and manual review of the following data : computer analysis using the program Genefinder (P. Green and L. Hallier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans Genome cloning project (<http://wofdb.dfci.harvard.edu/>), similarity to other proteins from BlastX analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES
source

```

1. 46000
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="II"
/clone="Y46E12BL"
4649. 8537
/gene="Y46E12BL.4"
join(4649. 5004,5059. 5096,6533. 6624,7842. 8537)
/gene="Y46E12BL.4"
/standard_name="Y46E12BL.4"
/notes="contains similarity to Pfam domain PF00622 (SPRY domain)"
/codon_start=1
/product="Hypothetical protein Y46E12BL.4"
/protein_id="AAK68490.1"
/db_xref="GI:14574357"
/db_xref="WormBase:Y46E12BL.4"
/translations="MDDFLPDRPDRIMQTKPDQIQEHSWNPEDRSINIFVKDED
KTFHHPVQAQSDTCIRGKMGYGFHVQIEWPERQGRGHVGVATKNAPLFAAGY
TALIGTDESIGMDISEPAVFGAKSEKKQRRCHDSKHTMTWYFNSRDVYVNP
DKFYCILDMDEGMATDDEDFGVAFRLNGKTLYPVIAAVWGHCIEINRWYLSLVR
ELISFSSSPTPRRLPPLPPSPSLQHHHRRSPPLPHSPVYKPKPSRTESPAL
KAAAAAASQSPVLPFPFSDLAITFSFYDLQAGSASSSSSYSLFPPPEHWFKRED
LPIDETPSASSPTTASLYLPAPSSSVRAANELFPGLTPGDSKITQDGLGLGTSI
L"
12457. 14434
/gene="Y46E12BL.3"
join(12457. 12619,13377. 13459,13838. 14003,14298. 14434)
/gene="Y46E12BL.3"
/standard_name="Y46E12BL.3"
/notes="contains similarity to Pfam domain PF00622 (SPRY domain); coded for by the following C. elegans cDNAs:
YK7862.3, YK7862.5"
/codon_start=1
/product="Hypothetical protein Y46E12BL.3"
/protein_id="AAK68490.1"
/db_xref="GI:14574356"
/db_xref="WormBase:Y46E12BL.3"
/translation="MGYSRGPHVQIEWPERQGRGHVGVATKNAPLQAEYTTING
SNSEYGMNIATRECHYGSWKYGPYNGRDGPNVPEKIYICILDMEOGNLSEATNEY
LGVVFQNLGRKLYPAVATLYGDCETIMTHLGSLEPVMPLPELFCRRRFRSIMGDDP
DKRIKPLMPPVYKVELLYEIE"
14971. 26567
/gene="Y46E12BL.2"
join(14971. 15038,15104. 15373,16296. 16369,16419. 16582,
17443. 17749,18910. 19586,21359. 22291,23583. 24883,
26357. 26567)
/gene="Y46E12BL.2"
/standard_name="Y46E12BL.2"
/notes="contains similarity to Interpro domains IPR000847

```

```

(Bacterial regulatory protein, LysR family), IPR002173
(carbohydrate kinase, PfkB); coded for by the following C.
elegans cDNAs: YK132e1.5, YK132e1.3, CEESH40F, YK193h8.5,
YK368e6.5, YK463b1.5, YK587d9.5, YK599b2.5, YK621g10.5"
/codon_start=1
/product="Hypothetical protein Y46E12BL.2"
/protein_id="AAK68489.1"
/db_xref="GI:14574355"
/db_xref="WormBase:Y46E12BL.2"
/translation="MGGNFRHRTSKGAKEFNSGACESNPTSKHRAANAARNAAS
FGAGMLAVEDRLMNTAAADMLLSALPIHSMDSKSSKSWSEGLSKLSGFTACTN
PAPSVNVRWKSQSSIOAEVSVLAAAEVIERKDEKDESVYCAALMTALEGSLGN
PRRTAYALLHLVRKPVKEVLAQHFRLVQIILYKLENNVDSGSALEKLSLVLG
ILRAQAQSVSSANTNMLSVLAALAHDKPVRTWAVRIRAVLTDVPTATDGLH
PASGAVQLINLHIESFLGRSSADNTNVRFLCLLEGVMHKMPAVLPKMAEAMLCF
AIDSVMKCSALQCLHRALORQCDLSALPTETNALLTLQLQASVTDVTVAYWQ
ALABHVCLTAKDSKSVQAFOFLPLFKIPEAGNEQLAQVTTQSILTRVENSIODD
AACGKFLISQLHVAMTKSASVMKFIILRAQMKUYETCGEQLGAEILTKWEDLARLQ
SDDCFCKEIDFTTGAARHIGVAVHVNILSLEVDPAAILSTFTSMLLPVLRVNI
HNAPISLIFISHFLPIAMKIHRLPTLAQVQVRLYSTFQFQWELLSPFCSPSDLET
AFPDIAPIGALNERNKDLMTVNLRRALRFLCPEADPAERVEVMRSRYAKNMFIF
FNMYTGSVDGDDKGVRLSLEAIRLYTEVTADLVTRYVDSALIAKSDHESVAS
QKQARVLIDCALKAENENSVKILDTIMWENSSTNGVQKAFILBEVIORRSH
PEIEHLDTENDYENALFKPIESISPARASYCSCVHMLDGCNTWKVEYDLTEKLI
RNIILLDKNNVHTRNSKCIQFICSKILDLVDPTIDTPSALEPVSLSKLYELTT
ISAEQKITGDIPLHVARSLVAAANIQAQVKIINTIATSKMISFACSWLLEGAAVR
ILAIRMLVKQKIPVLMQFQREQLITVFEQGLTSLDTIKVRKANRLLEVLVEKF
GVNVLQKTEKPDWLQMKNIERKKRKRKAGEAEMEHDEDEDGASAISSNVSGRTA
GADTILELLEDSDADNESDEEOLMKSRVGSVWLKNESGDVPMDLDSRMDMQIT
TMNPAILKPKKALERRQKVEAGFPQSGDKRMVLDGDEDEKQIGIKRRKRSYD
GLGDDDEDDGFTKGRKVENSDSEITLSDDDDDDEKDAKTFVSRAPSYGGKG
IYRDTSGKSVGSSVGAKRQKQKPNKQKQKSVQPYAYVPLRNKTAKQDVRQILKA
KRKSGKGGKKVTF"
complement(26739. 36107)
/gene="Y46E12BL.1"
complement(join(26739. 26980,29347. 29515,29922. 30193,
36026. 36107))
/gene="Y46E12BL.1"
/standard_name="Y46E12BL.1"
/notes="contains similarity to Pfam domain PF00019
(Transforming growth factor beta like domain)"
/codon_start=1
/product="Hypothetical protein Y46E12BL.1"
/protein_id="AAK68488.1"
/db_xref="GI:14574354"
/db_xref="WormBase:Y46E12BL.1"
/translation="MAMHELENSLIHCQEVTSILCRRLPSLGTCTLPSELVDCFYQ
SNSINHEILSASLIIDPKDTNISIVVEVDELFGELQVDFEIEITLTKYHFDISH
LFHKWQKQSKDKMIKIEITNSNTQNVNALSVLNAPNFVNFQFNTVGTGSDCV
GCCVPFYVNFTEIGNDWILSPGPFYANVCSDTVCGSTESDEVYQFMKAISDLPEPK
CAPNYGSDMTVALSPDRIRKRVHGLRALSCST"
BASE COUNT 14495 a 8606 c 7901 g 14998 t
ORIGIN
Alignment Scores:
Pred. No.: 63.8 Length: 46000
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 3 Gaps: 0
US-09-712-338-2 (1-555) x AC084156 (1-46000)
QY 501 AlaSerLeuGlnLeuPheAsnArgThrIle 510
|||||
Db 44495 GCTAGTCGCACTTTTAAAGAACATC 44524
|||||
RESULT 11
SG3789/c
LOCUS
DEFINITION
S63789 144 bp DNA linear PRI 01-SEP-1993
[SCZD1 marker, polymorphic microsatellite YN(CT)n] [human,
nucleated blood cells, genomic, 144 nt].

```



```

BASE COUNT      99 a   135 c   129 g   104 t
ORIGIN

Alignment Scores:
Pred. No.:      10.9      Length:      467
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    1.62%      Indels:      0
DB:             1      Gaps:      0

US-09-712-338-2 (1-555) x KSP302770 (1-467)

QY      88 GluAlaArgHisAsnProGluThrAla 96
Db      461 GAAGCAGCACACACCGGAGCGCA 435

RESULT 14
PPU296313/c
LOCUS      PPU296313      571 bp      DNA      linear      BCT 07-JUL-2002
DEFINITION Pseudomonas putida partial transposon Tn5053v4, tniQ-res-tniR
region.
ACCESSION      AJ296313
VERSION      AJ296313.1 GI:16200274
KEYWORDS      resolvase; tniQ gene; tniR gene; transposition protein TniQ;
transposon.
SOURCE      Pseudomonas putida
ORGANISM      Pseudomonas putida
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
AUTHORS      Mindlin,S., Kholodii,G., Gorlenko,Z., Minakhina,S., Minakhin,L.,
Kalyaeva,E., Kopteva,A., Petrova,M., Yurieva,O. and Nikiforov,V.
Mercury resistance transposons of gram-negative environmental
bacteria and their classification
JOURNAL      Res. Microbiol. 152 (9), 811-822 (2001)
MEDLINE      21604134
PUBMED      11763242
REFERENCE
AUTHORS      Kholodii,G.Y.
TITLE      Direct Submission
JOURNAL      Submitted (27-NOV-2000) Kholodii G.Y., Russian Academy of Sciences,
Institute of Molecular Genetics, Kurchatov Sq.2, Moscow 123182,
Russia, RUSSIA
FEATURES
source
1..571
/organism="Pseudomonas putida"
/mol_type="genomic DNA"
/strain="BW20"
/db_xref="taxon:303"
/country="USA:Bethesda"
/note="isolated from a stream"
1..571
/transposon="Tn5053v4"
complement(1..65)
/gene="tniR"
/gene="tniR"
/function="resolution of transpositional cointegrates"
/codon_start=1
/transl_table=11
/product="putative resolvase"
/protein_id="CAC82973.1"
/db_xref="GI:16200275"
/db_xref="SPTREMBL:Q936V4"
/translation="MLIGMRVSKADGSGQATDLQR"
73..102
/gene="tniR"
73..102
/gene="tniR"
/note="putative res subsite I, in which crossingover
occurs"
/bound_moiety="resolvase"

```

```

gene      complement(126..571)
/ gene="tniQ"
CDS      complement(126..>571)
/ gene="tniQ"
/ function="unknown"
/ codon_start=3
/ transl_table=11
/ product="putative transposition protein TniQ"
/ protein_id="CAC82974.1"
/ db_xref="GI:16200276"
/ db_xref="SPTREMBL:Q936V3"
/ translation="GQSLWRPPEYLAPAVQWQMLEAAATAITLIESKALIPRGQAAAL
EQEPHPTGFTNGLPKPKPEFPHNWRACAAIDEALIEARHNPETARLLFTLASVGR
RDPESLERLRATFTKGIPEFPEFWSHPEFWFAGRLNDGLSDSF"
BASE COUNT      103 a   176 c   171 g   121 t
ORIGIN

Alignment Scores:
Pred. No.:      13.1      Length:      571
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    1.62%      Indels:      0
DB:             1      Gaps:      0

US-09-712-338-2 (1-555) x PPU296313 (1-571)

QY      88 GluAlaArgHisAsnProGluThrAla 96
Db      323 GAAGCGGACACACCGGAGCGGCC 297

RESULT 15
AY143899
LOCUS      AY143899      1422 bp      mRNA      linear      PLN 05-OCT-2002
DEFINITION Arabidopsis thaliana At5g42240/K5J14_4 mRNA, complete cds.
ACCESSION      AY143899
VERSION      AY143899.1 GI:23505956
KEYWORDS      FLI CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 1422)
Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hayashizaki,Y.,
Hsuan,V.W., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
Nguyen,M., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
Southwick,A., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Arabidopsis ORF clones
Unpublished
2 (bases 1 to 1422)
Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hayashizaki,Y.,
Hsuan,V.W., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
Nguyen,M., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
Southwick,A., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (19-AUG-2002) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN
Arabidopsis Full-Length cDNA') : Seki,Y., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

```

COMMENT

The Salk, Stanford, PGSC (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAFL cDNAs: Cheuk, R., Chen, H., Kim, C.J., Shinn, P., Banh, J., Bowser, L., Chan, M.M., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Davis, R.W., Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES

source

Location/Qualifiers

1..1422
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="U12095"
/note="This clone is in PUNI 51"
ecotype: Columbia"

CDS

1..1422
/note="serine carboxypeptidase II-like"
/codon_start=1
/product="At5g42240/KSj14_4"
/protein_id="AAN28838.1"
/db_xref="GI:23505957"
/translation="MASVSRVAVAMVVVLLSLOWFAKGYPEEDLVVRLPGQPTVGF
KQYAGYVDVAVAGRSFLFYVEAVKQPSKELTLWNGPGCSSIGGGAFTLGPFF
PTGDGRGLRVNMSWNAKSHLLFVESPAGVWSYNSKSDYNTGDKSTANDMLVFLRL
WTEKPKLSRDLFTGESYAGHYIPQADAILSYNSHSSGPKFNKIGVAIGNPLIKL
DRSPATVEFFWSHGMI SDELKLTITSQCDDYTFASPHNVSTACNEAISETENIIT
EYNNYDVLDDVCYPSIVQGLRLKMKATKMSMGVDVCMVYERRFYENLPEVOKALHA
NRTHLPYSWMSGVLNYSIDIGNIDMLPILKRIILNKTPILWIFSGDDQSDVVPFGGSR
TLVRELAQDLNFKTTPYGAWFHKSQVGGWALEYKGLLIFATVRGAHVVPYAPQPSRA
LHLFSSFSVSGRLPNNTHSTDE"

BASE COUNT 373 a 303 c 351 g 395 t
ORIGIN

Alignment Scores:

Pred. No.:	30.6	Length:	1422
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.62%	Indels:	0
DB:	8	Gaps:	0

US-09-712-338-2 (1-555) x AY143899 (1-1422)

OY 99 ThrLeuTrpLeuAsnGlyGlyProGly 107

Db 232 ACCCTTGCTCAATGAGGTCAGGT 258

Search completed: November 21, 2003, 21:58:10
Job time : 4941 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 14:18:21 ; Search time 6062 Seconds
(without alignments)
11256.558 Million cell updates/sec

Title: US-09-712-338-1
Perfect score: 1668
Sequence: 1 atgcgtggtacgaattctt.....ccagtggtgtgcatgcatag 1668

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 2899711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_bt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1662	99.6	1662	6	AR129928	Sequence
2	533	32.0	2245	8	AF394242	Aspergill
3	25	1.5	1656	6	AX534871	Sequence
4	25	1.5	3150	6	AX534814	Sequence
5	21	1.3	1509	12	AF201927	Sequence
6	21	1.3	1539	6	AX306431	AF201927 Synthetic
7	21	1.3	128168	10	AL806526	AX306431 Sequence
8	20	1.2	3850	6	E12103	AL806526 Mouse DNA
9	20	1.2	3850	8	PPRC1GEN	E12103 DNA encodin
10	20	1.2	10029	1	AE011778	X87987 P.pastoris
11	20	1.2	10029	1	AE012241	AE011778 Xanthomon
12	20	1.2	11115	1	AE013499	AE012241 Xanthomon
13	20	1.2	45334	9	AC104595	AE013499 Methanosa
14	20	1.2	82843	9	AC104595	AC104595 Homo sapi
15	20	1.2	96161	8	AB062091	AP004005 Oryza sat
16	20	1.2	110000	2	AC127928_2	AB062091 Arabidops
17	20	1.2	11895	2	AC074108	Continuation (3 of
18	20	1.2	15041	3	CEY41E3	AC074108 Arabidops
19	20	1.2	154180	8	AP000399	Z95559 Caenorhabdi
20	20	1.2	161470	9	AC098851	AP000399 Oryza sat
21	20	1.2	171099	9	AC092644	AC098851 Homo sapi
22	20	1.2	172850	9	AC092644	AC092644 Homo sapi
23	20	1.2	178405	2	AC025927	AC092644 Homo sapi
24	20	1.2	194069	2	AC128533	AC025927 Homo sapi
25	20	1.2	235953	2	AC126554	AC128533 Rattus no
26	20	1.2	242069	2	AC126587	AC126554 Rattus no
27	20	1.2	242505	2	AC111849	AC126587 Rattus no
28	20	1.2	247314	2	AC123498	AC111849 Rattus no
29	20	1.2	248714	2	AC095710	AC123498 Rattus no
30	20	1.2	249633	2	AC127864	AC095710 Rattus no
31	20	1.2	251079	2	AC096217	AC127864 Rattus no
32	20	1.2	296950	1	AP001508	AC096217 Rattus no
33	20	1.2	311281	1	AE016859	AP001508 Bacillus
34	20	1.2	338579	1	AP003004	AE016859 Pseudomon
35	19	1.1	212	10	MUSLCA05	AP003004 Mesorhizo
36	19	1.1	744	10	MUSLY5AS	M23130 Mouse leuco
37	19	1.1	989	6	AX432827	M17320 Mus musculu
38	19	1.1	2876	4	AF163828	AX432827 Sequence
39	19	1.1	3470	4	AF163827	AF163828 Bos tauru
40	19	1.1	3475	4	AF163826	AF163827 Bos tauru
41	19	1.1	5007	10	MUSLY5AA	AF163826 Bos tauru
42	19	1.1	5220	1	AF479753	M29933 Mouse lymph
43	19	1.1	10529	1	AE005138	AF479753 Lactobaci
44	19	1.1	11208	1	HRNAPOF	AE005138 Halobacte
45	19	1.1	11502	1	AE011979	X57144 H. halobium
						AE011979 Xanthomon

ALIGNMENTS

RESULT 1
AR129928
LOCUS AR129928
DEFINITION Sequence 1 from patent US 6187578.
ACCESSION AR129928
VERSION AR129928.1 GI:14117825
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1662)
AUTHORS Blinksy, A., Berka, R., Golightly, E., Klotz, A.,
Mathisen, T. Erik., Dambmann, C. and Brown, K.M.
TITLE Carboxypeptidases and nucleic acids encoding the same
JOURNAL Patent: US 6187578-A 1 13-FEB-2001;

FEATURES		Location/Qualifiers	
source		1..1662	
BASE COUNT		396 a 468 c 398 g 400 t	
ORIGIN			
Query Match		99.6%; Score 1662; DB 6; Length 1662;	
Best Local Similarity		100.0%; Pred. No. 0;	
Matches 1662; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	ATGCGTGGCTACGAAATTTCTCTAGTGTACCTTGGTTGCGAGCCAGTTCGGCCCTTCCA	60
Db	1	ATGCGTGGCTACGAAATTTCTCTAGTGTACCTTGGTTGCGAGCCAGTTCGGCCCTTCCA	60
Qy	61	GGAAGTACACCGCGTTCGTCGGTAGAAGACAGTACCCAAAGAACCCCGGGGTCAAG	120
Db	61	GGAAGTACACCGCGTTCGTCGGTAGAAGACAGTACCCAAAGAACCCCGGGGTCAAG	120
Qy	121	ACTCTTAAACCGGAAACAAATGTCCATCCGGTACAAAGAACCCGGGGCAGAGGGCTC	180
Db	121	ACTCTTAAACCGGAAACAAATGTCCATCCGGTACAAAGAACCCGGGGCAGAGGGCTC	180
Qy	181	TGCAGACTACCCGGGTGTCAATCTTACTCTGGATATGCGACACCTCTCCGAGTCC	240
Db	181	TGCAGACTACCCGGGTGTCAATCTTACTCTGGATATGCGACACCTCTCCGAGTCC	240
Qy	241	CATACCTTCTTCTGTTCTTTCGAAGCCAGACATACCCAGAAACTGCACTTATCACATTG	300
Db	241	CATACCTTCTTCTGTTCTTTCGAAGCCAGACATACCCAGAAACTGCACTTATCACATTG	300
Qy	301	TGGTTGAATGGTGGCCCTGGAAGCGANTTTTGTATCGGTCTCTTGAAGAGTTGGGCCCT	360
Db	301	TGGTTGAATGGTGGCCCTGGAAGCGANTTTTGTATCGGTCTCTTGAAGAGTTGGGCCCT	360
Qy	361	TGCCATGTCAATTCACATTTTATCAGTACATCAACCTCTACTCTGTAAGAGAGTTC	420
Db	361	TGCCATGTCAATTCACATTTTATCAGTACATCAACCTCTACTCTGTAAGAGAGTTC	420
Qy	421	AATTTTACTATTCTCTCCAGCAATGGGAGTTCGGCTTTTTCATATAGTATGATACGTTGAT	480
Db	421	AATTTTACTATTCTCTCCAGCAATGGGAGTTCGGCTTTTTCATATAGTATGATACGTTGAT	480
Qy	481	GGGTCCATTAACCTCTGTAACCTGGGTGCTCGAATAATTCGAGTTTTCGAGGATTCAGGC	540
Db	481	GGGTCCATTAACCTCTGTAACCTGGGTGCTCGAATAATTCGAGTTTTCGAGGATTCAGGC	540
Qy	541	CGGTACCAACCAATTTGATGCGCACTCTGATCGATACCTACCAATTCGCGCAGAGCCGCT	600
Db	541	CGGTACCAACCAATTTGATGCGCACTCTGATCGATACCTACCAATTCGCGCAGAGCCGCT	600
Qy	601	TGGGAGATCTCTGCAAGGATTTCTTAGTGGACTTACCTAGTTCGACTTAGGGTGCAGTCT	660
Db	601	TGGGAGATCTCTGCAAGGATTTCTTAGTGGACTTACCTAGTTCGACTTAGGGTGCAGTCT	660
Qy	661	AAGGCTTCAGTCTATGACGAGAGACTATGAGGGCACTATGGTCTGCTCAATTCCTCAAT	720
Db	661	AAGGCTTCAGTCTATGACGAGAGACTATGAGGGCACTATGGTCTGCTCAATTCCTCAAT	720
Qy	721	CATTTTACGACGAGATCAGAGAAATTCGAACGGTGTGTAATGGTCTTCAGCTTAAT	780
Db	721	CATTTTACGACGAGATCAGAGAAATTCGAACGGTGTGTAATGGTCTTCAGCTTAAT	780
Qy	781	TTCAACTCTCTGGGAATTTAAGCGCATCATCGACGAGCGCATCCAGGCCCTTACTAC	840
Db	781	TTCAACTCTCTGGGAATTTAAGCGCATCATCGACGAGCGCATCCAGGCCCTTACTAC	840
Qy	841	CCTGAATTCGCTGTGACAACTACCTACGTTATCAAGGCTGTCAACGAGACCGCTTACAAC	900
Db	841	CCTGAATTCGCTGTGACAACTACCTACGTTATCAAGGCTGTCAACGAGACCGCTTACAAC	900
Qy	901	TACATGAAGTTTCCCAACCAAAATGCGAAATGGTTGCGAGATTTGATTTCCACCTCAAA	960
Db	901	TACATGAAGTTTCCCAACCAAAATGCGAAATGGTTGCGAGATTTGATTTCCACCTCAAA	960

RESULT 2

AF394242

LOCUS

DEFINITION

Aspergillus oryzae strain TK3 carboxypeptidase S1 (cpi) gene,

complete cds.

ACCESSION

AF394242

VERSION

AF394242.1

KEYWORDS

GT:15004615

SOURCE

ORGANISM

Aspergillus oryzae

Aspergillus oryzae

2245 bp DNA linear

PLN 24-JUL-2001

Aspergillus oryzae strain TK3 carboxypeptidase S1 (cpi) gene,

complete cds.

AF394242

GT:15004615

Aspergillus oryzae

Aspergillus oryzae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

van den Broek, P.

Direct Submission

Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,

Lausanne CH-1000, Switzerland

Location/Qualifiers

1..2245

Aspergillus oryzae

Aspergillus oryzae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

van den Broek, P.

Direct Submission

Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,

Lausanne CH-1000, Switzerland

Location/Qualifiers

1..2245

/organism="Aspergillus oryzae"

/mol_type="genomic DNA"

/strain="TK3"

/db_xref="taxon:5062"

```
gene
<1..>2245
/gene="cpi"
join(<1..349,409..513,576..692,743..866,941..956,
1008..1124,1187..1237,1296..1449,1522..1580,1670..2245)
/gene="cpi"
/product="serine carboxypeptidase 1"
join(<1..349,409..513,576..692,743..866,941..956,
1008..1124,1187..1237,1296..1449,1522..1580,1670..2245)
/gene="cpi"
/EC number="3.4.16.6"
/codon_start=1
/product="carboxypeptidase S1"
/protein_id="AAK7166.1"
/db_xref="GI:115004616"
/translation="MRGYFLSLVPLVAASWALPGSTPASVGRQLPKNPTGVKTLTT
ANNVIRYKPGAGVCECTPGVKSQVVDTSPEHTEFWFPEARHNPETAITLWL
NGGPGSDSLGLPELGPCHVNSTFDYINPHSNVSNLLFLSQBLGFGFSYNDTVD
GSINPTGVVNSFSPAGVQGRYPTIDATIDITNLAAEAWEILQFLSGLSLDSRV
QKDFSLWTSYGRHGPAPFFHFYEQNRIANGSVQNLNFSGLIINGIDEAIO
APYTFPANNVTYIKAVNETVYNYMKFANQMGQDILSTCKQNRNALADYALCA
EATNCRDNVEGYPYAFAGRVYDIRHPYDDPTPPSYNKLAKDSVMDAIGVINVT
QSNNDVYAFQQTGDFWPNFIEDLEILALPVRSLYIGDADYICNWFQGAQSLAA
NYSQAQFSAGVTLPLKNGVVEGETREYCNESFTRYEAGHEVPYQPIASLQLENR
TIFGWDIABGQKIMSYKINGTATATHTIQSSVPLETATSMSSVGNVA"
sig_peptide
1..54
/gene="cpi"
BASE COUNT 548 a 603 c 524 g 570 t
ORIGIN

Query Match 32.0%; Score 533; DB 8; Length 2245;
Best Local Similarity 100.0%; Pred. No. 1.4e-302;
Matches 533; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 AAGGACTCTGTATGACGCTATCGCGCTCAACATCAATACACCCAGTCCATATGAC 1191
DB 1709 AAGGACTCTGTATGACGCTATCGCGCTCAACATCAATACACCCAGTCCATATGAC 1768
QY 1192 GTCTACTACGTTTCAGCAACAGCGCACTTTGTCTGGCCCACTTCATCGAAGACCTC 1251
DB 1769 GTCTACTACGTTTCAGCAACAGCGCACTTTGTCTGGCCCACTTCATCGAAGACCTC 1828
QY 1252 GAGGAGATCTTCTCTCCCGTGGTGTCTCCCTCATCTATGGGACCGCATACATC 1311
DB 1829 GAGGAGATCTTCTCTCCCGTGGTGTCTCCCTCATCTATGGGACCGCATACATC 1888
QY 1312 TGCAACTGTTCCGGCGGTGAGCGGCTTCCCTCGCTGCAATCTCCCAAGCCGCCAG 1371
DB 1889 TGCAACTGTTCCGGCGGTGAGCGGCTTCCCTCGCTGCAATCTCCCAAGCCGCCAG 1948
QY 1372 TTCGAAGCGGAGGTACACGCCCTCGAAGTCAACGGCGTGCAGTATGGGAAACTCGC 1431
DB 1949 TTCGAAGCGGAGGTACACGCCCTCGAAGTCAACGGCGTGCAGTATGGGAAACTCGC 2008
QY 1432 GAGTATGTAATTTCTCTCTCACTCGCTGCTATGAGGAGGCGCATGAAGTCCCATAC 1491
DB 2009 GAGTATGTAATTTCTCTCTCACTCGCTGCTATGAGGAGGCGCATGAAGTCCCATAC 2068
QY 1492 CAGCCCATCGCTCCCTCGCAATTTTAACCGGACTATCTCGGTGGGATATCCGACAG 1551
DB 2069 CAGCCCATCGCTCCCTCGCAATTTTAACCGGACTATCTCGGTGGGATATCCGACAG 2128
QY 1552 GGCAGAGAGAGATCTGCCCGAGCTACAAGAGCAATGGAACGGCTACAGCTACGATACA 1611
DB 2129 GGCAGAGAGAGATCTGCCCGAGCTACAAGAGCAATGGAACGGCTACAGCTACGATACA 2188
QY 1612 CAGTGTCCGTCGCCCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCT 1664
DB 2189 CAGTGTCCGTCGCCCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCT 2241

RESULT 3
AX534871
LOCUS
DEFINITION Synthetic construct gag protein gene, complete cds.
AF201927
LOCUS 1509 bp DNA linear SYN 16-MAR-2000
```

```
ACCESSION      AF201927
VERSION        AF201927.1  GI:7248702
KEYWORDS       synthetic construct
SOURCE         synthetic construct
ORGANISM       artificial sequences.
REFERENCE      1 (bases 1 to 1509)
AUTHORS        zur Megede,J., Chen,M.C., Doe,B., Schaefer,M., Greer,C.E.,
                Selby,M., Otten,G.R. and Barnett,S.W.
TITLE          Increased expression and immunogenicity of sequence-modified human
                immunodeficiency virus type 1 gag gene
JOURNAL        J. Virol. 74 (6), 2628-2635 (2000)
MEDLINE        20148954
PUBMED         10684277
REFERENCE      2 (bases 1 to 1509)
AUTHORS        zur Megede,J. and Barnett,S.W.
TITLE          Direct Submission
JOURNAL        Submitted (04-NOV-1999) Vaccines, Chiron Corporation, 4560 Horton,
                Emeryville, CA 94609, USA
FEATURES       Location/Qualifiers
                source
                1..1509
                /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
                1..1509
                /note="HIV-1SF2 p55gag"
                /codon_start=1
                /transl_table=1
                /product="gag protein"
                /protein_id="AAP43628.1"
                /db_xref="GI:7248703"
                /translation="MGARASVLSGELDKWEKIRLPQGGKKYKHLKHIVWASRELERF
                AVNPLGILTSRGQILQLOPISLTQSEELRSYNTVATLYCVHQRIDVDKDTKEALE
                KTEBQNSKKKAQQAAGATGNSQVSNYPVONLQGWVHOAIPRTLNWVKV
                VEKAPSEVTPMSALSEGATPDQDLNMLNTVGGHQAMQMLKETINEEAENDRVH
                PVHGIAPQGMREPRGSDIAGTISTLQEQIGWNNPPIPVGEIYKXWILGLNKIV
                RMYSTSLDRQGEKPEFRDYVDYFKTLRAEQASQDVKNWNTTLLVQNPDPCKT
                ILKALGPAATLEEMTACQGGVGGFKARVLAEMSQVTNATIMVQRGNRNQKTV
                KFCNGKEGHTARNRCKRKGCGREGHQMDCTERQANFLGKIWPYSYKRGNGF
                LQSRPEPTAPPEPSFRFGEEKTPSKQEPIDKELYPLTSLRSLFGNDPSSQ"
BASE COUNT    328 a 543 c 479 g 159 t
ORIGIN
Query Match      1.3%; Score 21; DB 12; Length 1509;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCGAGCTACAAG 1581
      |||||
DB 1312 AAGATCTGGCCGAGCTACAAG 1332

RESULT 6
AX306431
LOCUS          AX306431 1539 bp DNA linear PAT 11-DEC-2001
DEFINITION    Sequence 5 from Patent WO0188141.
ACCESSION     AX306431
VERSION       AX306431.1 GI:17645655
KEYWORDS      synthetic construct
SOURCE        synthetic construct
ORGANISM      artificial sequences.
REFERENCE      1
AUTHORS        Wagner,R., Graf,M., Deml,L. and Bieler,K.
TITLE          Synthetic gagpol genes and their uses
JOURNAL        Patent: WO 0188141-A 5 22-NOV-2001;
                Genearth GmbH (DE)
FEATURES       Location/Qualifiers
                source
                1..1539
                /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
                /note="sequence with optimized codons"

ACCESSION      AF201927
VERSION        AF201927.1  GI:7248702
KEYWORDS       synthetic construct
SOURCE         synthetic construct
ORGANISM       artificial sequences.
REFERENCE      1 (bases 1 to 1509)
AUTHORS        zur Megede,J., Chen,M.C., Doe,B., Schaefer,M., Greer,C.E.,
                Selby,M., Otten,G.R. and Barnett,S.W.
TITLE          Increased expression and immunogenicity of sequence-modified human
                immunodeficiency virus type 1 gag gene
JOURNAL        J. Virol. 74 (6), 2628-2635 (2000)
MEDLINE        20148954
PUBMED         10684277
REFERENCE      2 (bases 1 to 1509)
AUTHORS        zur Megede,J. and Barnett,S.W.
TITLE          Direct Submission
JOURNAL        Submitted (04-NOV-1999) Vaccines, Chiron Corporation, 4560 Horton,
                Emeryville, CA 94609, USA
FEATURES       Location/Qualifiers
                source
                1..1509
                /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
                1..1509
                /note="HIV-1SF2 p55gag"
                /codon_start=1
                /transl_table=1
                /product="gag protein"
                /protein_id="AAP43628.1"
                /db_xref="GI:7248703"
                /translation="MGARASVLSGELDKWEKIRLPQGGKKYKHLKHIVWASRELERF
                AVNPLGILTSRGQILQLOPISLTQSEELRSYNTVATLYCVHQRIDVDKDTKEALE
                KTEBQNSKKKAQQAAGATGNSQVSNYPVONLQGWVHOAIPRTLNWVKV
                VEKAPSEVTPMSALSEGATPDQDLNMLNTVGGHQAMQMLKETINEEAENDRVH
                PVHGIAPQGMREPRGSDIAGTISTLQEQIGWNNPPIPVGEIYKXWILGLNKIV
                RMYSTSLDRQGEKPEFRDYVDYFKTLRAEQASQDVKNWNTTLLVQNPDPCKT
                ILKALGPAATLEEMTACQGGVGGFKARVLAEMSQVTNATIMVQRGNRNQKTV
                KFCNGKEGHTARNRCKRKGCGREGHQMDCTERQANFLGKIWPYSYKRGNGF
                LQSRPEPTAPPEPSFRFGEEKTPSKQEPIDKELYPLTSLRSLFGNDPSSQ"
BASE COUNT    328 a 543 c 479 g 159 t
ORIGIN
Query Match      1.3%; Score 21; DB 12; Length 1509;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCGAGCTACAAG 1581
      |||||
DB 1312 AAGATCTGGCCGAGCTACAAG 1332

RESULT 6
AX306431
LOCUS          AX306431 1539 bp DNA linear PAT 11-DEC-2001
DEFINITION    Sequence 5 from Patent WO0188141.
ACCESSION     AX306431
VERSION       AX306431.1 GI:17645655
KEYWORDS      synthetic construct
SOURCE        synthetic construct
ORGANISM      artificial sequences.
REFERENCE      1
AUTHORS        Wagner,R., Graf,M., Deml,L. and Bieler,K.
TITLE          Synthetic gagpol genes and their uses
JOURNAL        Patent: WO 0188141-A 5 22-NOV-2001;
                Genearth GmbH (DE)
FEATURES       Location/Qualifiers
                source
                1..1539
                /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
                /note="sequence with optimized codons"
```

```
BASE COUNT    366 a 508 c 502 g 163 t
ORIGIN
Query Match      1.3%; Score 21; DB 6; Length 1539;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCGAGCTACAAG 1581
      |||||
DB 1306 AAGATCTGGCCGAGCTACAAG 1326

RESULT 7
AL806526
LOCUS          AL806526 128168 bp DNA linear ROD 09-AUG-2002
DEFINITION    Mouse DNA sequence from clone RP23-123F24 on chromosome 11,
                complete sequence.
ACCESSION     AL806526
VERSION       AL806526.4 GI:22204685
KEYWORDS      HTG.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 128168)
AUTHORS        Kay,M.
TITLE          Direct Submission
JOURNAL        Submitted (01-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
                Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
                On Aug 11, 2002 this sequence version replaced gi:21912768.
COMMENT        ----- Genome Center
                Center: Wellcome Trust Sanger Institute
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquery@sanger.ac.uk
                -----
                During sequence assembly data is compared from overlapping clones.
                Where differences are found these are annotated as variations
                together with a note of the overlapping clone name. Note that the
                variation annotation may not be found in the sequence submission
                corresponding to the overlapping clone, as we submit sequences with
                only a small overlap as described above.
                This sequence was finished as follows unless otherwise noted: all
                regions were either double-stranded or sequenced with an alternate
                chemistry or covered by high quality data (i.e., phred quality >=
                30); an attempt was made to resolve all sequencing problems, such
                as compressions and repeats; all regions were covered by at least
                one plasmid subclone or more than one M13 subclone; and the
                assembly was confirmed by restriction digest. The following
                abbreviations are used to associate primary accession numbers given
                in the feature table with their source databases: Em.; EMBL; SW.;
                SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP
                database can be found at
                http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-123F24 is
                from the RPI-23 Mouse PAC library
                constructed by the group of Pieter de Jong.
                For further details see http://www.chori.org/bacpac/home.htm
                VECTOR: pBACE3.6.

FEATURES       Location/Qualifiers
                source
                1..128168
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10090"
                /chromosome="11"
                /clone="RP23-123F24"
                /clone_lib="RPI-23"
                /clone_lib="RPI-23"

BASE COUNT    36202 a 29519 c 28222 g 34225 t
ORIGIN
Query Match      1.3%; Score 21; DB 10; Length 128168;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


QY 1465 GAGCAGGCCATGAAGTCCCA 1485
Db 9958 GAGCAGGCCATGAAGTCCCA 9978

RESULT 8
E12103
LOCUS DNA encoding precursor of protease from Pichia pastoris. linear PAT 27-APR-1998
DEFINITION E12103
ACCESSION E12103
VERSION E12103.1 GI:3250937
KEYWORDS JP 1996256770-A/1.
SOURCE Pichia pastoris
ORGANISM Pichia pastoris
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE 1 (bases 1 to 3850)
AUTHORS Ohtani, H., Murakami, K., Okazaki, N., Ooya, T. and Omura, T.
TITLE PROTEASE DERIVED FROM YEAST BELONGING TO THE GENUS PICHIA
JOURNAL Patent: JP 1996256770-A 1 08-OCT-1996;
GREEN CROSS CORP:THE

COMMENT OS Pichia pastoris
PN JP 1996256770-A/1
PD 08-OCT-1996
PF 24-MAR-1995 JP 1995066597
PI OI HIDEYUKI, OTANI WATARU, MURAKAMI KOJI, OKAZAKI NORIKO, PI
OOYA TOMOSUKE,
PI OMURA TAKAO
PC C12N9/60, C07H21/04, C12N1/19, C12N15/09, C12P21/02, (C12N9/60, PC
C12R1:84),
PC (C12N1/19, C12R1:84), (C12P21/02, C12R1:84);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..3850
FT /organism="Pichia pastoris"
FT /strain="GTS115"
FT 5'UTR 1..2064
FT sig_peptide 2065..2124
FT misc_feature 2125..2385
FT /notes="Propeptide region of precursor of FT
protease from
Pichia pastoris"
FT mat_peptide 2386..3636
FT /product="Protease from Pichia pastoris" FT
3'UTR 3637..3850
FT misc_feature 2014..2019
FT /notes="like TATA box".

FEATURES
source
Location/Qualifiers
1..3850
/organism="Pichia pastoris"
/mol_type="genomic DNA"
/db_xref="taxon:4922"
a 760 c 760 g 1145 t

BASE COUNT 1185 a 760 c 760 g 1145 t
ORIGIN

Query Match 1..2%; Score 20; DB 6; Length 3850;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 TGGTTGAATGGTGGCCCTGG 320
Db 2527 TGGTTGAATGGTGGCCCTGG 2546

RESULT 9
PPRC1GEN
LOCUS P. pastoris PRC1 gene. linear PLN 17-AUG-1998
DEFINITION X87987
ACCESSION X87987.1 GI:1171615
VERSION

KEYWORDS carboxypeptidase Y; PRC1 gene.
SOURCE Pichia pastoris
ORGANISM Pichia pastoris
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE 1
AUTHORS Ohtani, H., Okazaki, N., Furuhashi, N. and Ohmura, T.
TITLE Cloning and characterization of the Pichia pastoris PRC1 gene
JOURNAL encoding carboxypeptidase Y
MEDLINE Yeast 12 (1), 31-40 (1996)
96381245
PUBMED 8789258
REFERENCE 2 (bases 1 to 3850)
AUTHORS Ohtani, H.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1995) H. Ohtani, Central Research Laboratory, The
Green Cross Corporation, 2-25-1 Shodai-Ohtani, Hiraka, Osaka 573,
JAPAN

FEATURES
source
Location/Qualifiers
1..3850
/organism="Pichia pastoris"
/mol_type="genomic DNA"
/strain="GTS115"
/db_xref="taxon:4922"
2014..2019
2065..3636
/gene="PRC1"
2065..3636
/gene="PRC1"
/codon_start=1
/product="carboxypeptidase Y"
/protein_id="CAA61240.1"
/db_xref="GI:1171615"
/translation="MILHTYIILSLTTFPKAIGLSLOMPKALEASVSLYEKATLAV
QOEIDAIQKIQGQWLEVEFPTIVSOLSYSTGPKPAKKKADTFWDFYVESQLFNP
YLRVKNNPEVLKVDFTKQYSGYLDVEADKHFYWFESRNPQNDPFIILLNGGP
CCSLTGLFFELGSRINENKPIFNPSYMNAGSIYLDQFVNVGFSYSSSVNTV
VAGEDVAFQLQFFQHPPEYQTFNDFHAGESYAGHYIPVFADEILSKNRNFTSVL
INGLTDLTQTYRYEPMACGEGAPSVLPADCECNMLVTQDKLSLIQACYSQSAF
TCAPAAIYCNNAQMPYQRTGKGVYDIRKECDGSLCYKLELEIDTVLNOKFQDLAG
AEVTYSCNFEINRNLPAQDMKPYHEHVSLLNKGSLPULIYVAGDKDFICNWLGNR
ANTQVLPVADADGEKAEVQDMLVNGRAGEFKNISNFTILRVYDAGHMAPYDQFENS
HEMVRNWSIGDFSPH"
2065..2124
/gene="PRC1"
2386..3633
/gene="PRC1"
/product="unnamed"
BASE COUNT 1185 a 760 c 760 g 1145 t
ORIGIN

Query Match 1..2%; Score 20; DB 8; Length 3850;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 TGGTTGAATGGTGGCCCTGG 320
Db 2527 TGGTTGAATGGTGGCCCTGG 2546

RESULT 10
AE011778/c
LOCUS Xanthomonas axonopodis pv. citri str. 306, section 156 of 469 of
DEFINITION the complete genome.
ACCESSION AE011778 AE008923
VERSION AE011778.1 GI:21107629
KEYWORDS Xanthomonas axonopodis pv. citri str. 306
SOURCE Xanthomonas axonopodis pv. citri str. 306
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.

REFERENCE
AUTHORS

1 (bases 1 to 10029)
da Silva A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Canarotte, G., Cannavan, F., Cardoso, J.,
Chamberg, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
Rossa, J.A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezra, R.I.D., Trindade dos
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
Kitajima, J.P.

TITLE

Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities

JOURNAL

Nature 417 (6887), 459-463 (2002)

MEDLINE

20022145

PUBMED

12024217

REFERENCE
AUTHORS

2 (bases 1 to 10029)
da Silva A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Canarotte, G., Cannavan, F., Cardoso, J.,
Chamberg, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
Rossa, J.A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezra, R.I.D., Trindade dos
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
Kitajima, J.P.

TITLE

Direct Submission

JOURNAL

Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil

FEATURES
source

Location/Qualifiers

1. .10029

/organism="Xanthomonas axonopodis pv. citri str. 306"

/mol_type="genomic DNA"

/strain="306"

/db_xref="taxon:130486"

/note="pathovar: citri"

218. .703

/gene="gpo"

/note="synonym: XAC1457"

218. .703

/gene="gpo"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/Genemark"

/codon_start=1

/transl_table=11

/product="glutathione peroxidase-like protein"

/protein_id="XAC36327.1"

/db_xref="GI:21107630"

/translation="MSDSTFAFTLEGRQTSRMRYAGNVLVNVNASKGFTPOYAG

LCALWQRGLVIGFPCDQGHQEGDAQIRQFSLDYAVDFPLAAKIDVNGNG

AHPLWQWLKHEQRGVLSGEAIKWNFTKIGRDLGRVRYAPTPTKPEALATEIERALD

A"

complement (854. .1633)

/gene="fpr"

/note="synonym: XAC1458"

complement (854. .1633)

/gene="fpr"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/Genemark"

/codon_start=1
/transl_table=11
/product="ferridoxin-NADP reductase"
/protein_id="XAC36328.1"
/db_xref="GI:21107631"
/translation="MSSARGAETVLEVRHWTDAYFSFTTRDAGRFEFGQFWMGLE
TETPLRAISIASANWEHEFFSIVKDPGLISLQHLQPGDKVIVGKPGTGLLI
SDLHGRNLYLGTGLAPWISIIKDPETFERFVKVLTQGVRFQVDLAYRDYPERE
LPQHEFGLDLREKLLYPAVTRFETANQGRTELMADGRMQQTGLTLPDANDRFM
ICSPQWLADRLSLDSRGFQTSRPTGCHYVFERAFVEK"
complement (1849. .3729)
/gene="msbA"
/note="synonym: XAC1459"
complement (1849. .3729)
/gene="msbA"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="XAC36329.1"
/db_xref="GI:21107632"
/translation="MSASRGATAPAAPRTGPTLRERLDALRNLPPFLOQIWTQSRWL
SASSIGLVRALMPVATLYIGKLIIDEATHLVGPSAPQFAQLGSGRLNLELL
ALBELAGLSDLLGRVYADALLSELFNAAVSRLMEHAQDLDFEDPDQDKLD
RARQTNRNWNLMSQLFGQVQDALTIVSLAGLVVAPWLVLLVALIIPAFGEAHE
NALGYSINFQTPFEROLDYRQGVASVETAKEVKILNHRFLIRRRADKRFQAN
RALAKRAMGALLAALGTLYGMAYGIATRTVGDFSGDITFLAGSLRLRLQLE
GLTIGFQVAGQALYLDLXSFNVIPEIRSNALPVPPIRQGFVFEDGFRYFDA
EOWTVRLHDLRAGEVLALVGENGAGKTLVKILARLYDPDEGRILLDQDLRDYDL
DELRANGLVIFQDPVRVHLSVGENIGVQVDMDDTARIRAAQRAVAGSLDLSDFK
YNQILGRFNSVDLSGGQWQALAIAPYWRDAQMLTDLPFTAALDARSSEFEVFPK
ELSDNRITAVLISHRFSSVRWADRLVLAAGRIEANGTHEELMAQGGRYAEFLQAAQ
YR"
4027. .5544
/gene="fumb"
/note="synonym: XAC1460"
4027. .5544
/gene="fumb"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="fumarate hydratase"
/protein_id="XAC36330.1"
/db_xref="GI:21107633"
/translation="WTSIKQEDLIQSVADALQYISYHVPDIYIKLSAAYERESPAA
KDAIAQILINSRMCAGEHRPICQDTGIVTVFLEIGMNVRRDDATLGVDMVNEGVRRA
YNHPDNKLRAVLADPAGKTRANTDTAVVTKIVPGDHVEIVAAKGGSEAKSKF
AMAPSDIVDMVLKVTMTGACWCPGMLGIGGTGAEKAMLLAKEALMBPIDIVDL
QARQSPNRAEELRELYEKVNALGIGAGGLTIVLDIKVQDYPHAAANLPVALIPN
CAATRHAAFTLDGGSPVMDLPPSLDWPGLTVPFNARRVNLDTISKEVASPKGET
ILLNGLKTGRDAHKKIMDLNREGTILPVDITNRFIYVGPVDPDRVVGPAFTT
ATRMDFKTRQMLEQTGLLGKSGERGDAIAARDNKAIVMAYMGVSAYLVSKAKA
ARVLAFEDLIGMEALVEVEKDPMTVAVDSSGESVHKTGPRQMSQRIKIPVVE"
5798. .6421
/gene="gst"
/note="synonym: XAC1461"
5798. .6421
/gene="gst"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="glutathione S-transferase"
/protein_id="XAC36331.1"
/db_xref="GI:21107634"
/translation="WHSPTAALVVMILLIETDIPHALHPLDPERGHEKSPQYLALN
PAGVYPTVLDQVLTAAIVLOIADLHPQAGLAPVGTPERAAYIKWMTFCANTLQ
PAYRAWFTSPHAPAGQBAQAQHAARKLEAAWTVQDAHLOAHGPLYLTKTVSAADF
ITMLNWRSNMRPDTITPALKAHANQMKARPAFQETVRRSLTDTW"
6456. .7247
/gene="cynT"

```

CDS
/notes="synonym: XAC1462"
6456..7247
/genes="cynT"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="carbonic anhydrase"
/protein_id="AAM36332.1"
/db_xref="GI:21107635"
/translation="MPLRVGSRPHSTIATPTTQGGSMHSGARRKLLQACACAP
IAAASLSPALRAEAKSKTDLNAEQALELRDGNFVQNRPRKVISDSKRLELAL
GQTFVLVSDSRVPELLFGRGEMFIVRNAGNTVDITLALGSIYAVSGLGVPL
VVMHGESGCAVAAVSVENNTFFPCAIGSMIEPIVPAVLAKAGGTDLLADSVKA
NVRTVDRLRAASEPALLEPLRAGNVVVGAYYTLQDGKVDFFDV"
complement(7610..8827)
/genes="XAC1463"
complement(7610..8827)
/genes="XAC1463"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="phospholipase"
/protein_id="AAM36333.1"
/db_xref="GI:21107636"
/translation="MYTLRMOTRPLSCASLRPLGAPTSARLPMSKHTEPLMLIALA
APLAQETAPQAPSPQACTSVSDAARLACYDKALGYTPQGTAEADAQAQAKQEA
KPDPTRAISRGDFRADQDQPKBEAVANAGRGLDRRLAKDSKIGTFLRGYK
FVYLLPAFWISDTNTPQSPNPANSVNTQQDLSTELQKLFKTKVWEDLFDNGDI
MGYTSQRQWYNITDASPRETNYEPAMLLFRNNYSIFGKMGSGISLNHMSNG
RADPLSRNRYILNFDRENWALTLEWFRFKEDRADDDNPDIEDVYMGCDATLVY
NKDGEFALIAHSLRGLGDRSHGSQLDYGFPIITLLRHGVQVDFDYGESMDYNHKA
TVIGVGSLLWF"
complement(9001..9486)
/genes="XAC1464"
complement(9001..9486)
/genes="XAC1464"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM36334.1"
Query Match 1-2%; Score 20; DB 1; Length 10029;
Best Local Similarity 100.0%; Pred.No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 808 ATCATCGACGAGCGATCCA 827
Db 3525 ATCATCGACGAGCGATCCA 3506
|||||
|||||

RESULT 11
AE012241/c
LOCUS
DEFINITION
Xanthomonas campestris pv. campestris str. ATCC 33913, section 149
of 460 of the complete genome.
ACCESSION
AE012241 AE008922
VERSION
AE012241.1 GI:21112475
KEYWORDS
Xanthomonas campestris pv. campestris str. ATCC 33913
Xanthomonas campestris pv. campestris str. ATCC 33913
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
SOURCE
1 (bases 1 to 10029)
Xanthomonas campestris pv. campestris str. ATCC 33913
ORGANISM
Xanthomonas campestris pv. campestris str. ATCC 33913
REFERENCE
AUTHORS
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira,Jr.,H.A.,
Rossa,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
Taktai,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kitajima,J.P.
Direct Submission
Submitted (28-NOV-2001) Departamento de Bioquimica, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 749, Sao Paulo, SP 05508-900,
Brazil
FEATURES
Location/Qualifiers
1..10029
/organism="Xanthomonas campestris pv. campestris str. ATCC
33913"
/mol_type="genomic DNA"
/strain="ATCC 33913"
/db_xref="ATCC:33913"
/db_xref="taxon:190485"
/notes="pathovar: campestris"
complement(145..2295)
/genes="dcp"
/notes="synonym: XCC1412"
complement(145..2295)
/genes="dcp"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="peptidyl-dipeptidase"
/protein_id="AAM40709.1"
/db_xref="GI:21112476"
/translation="MPSYSIAAPAAQAATANPFFADSTLPLHYPOFDKIDSDPAP
AFDAGMAQLKDEVAIANQAKPTFDNTIIALEKSGATLDRATTVFENLGVADTNDAR
KKLOADYSKFAAHRDAISLNGKLFARITQTHDQRSKGLDQAQVRLVEKYISDFVRD
GAKLSADAKTKLKMNAELANLTTFFSQNLVAEVNAAVVDDVQDLGDSQIAAA
AEAARKLDGKYVIALNTTGPPLTQKLRRLRRIYDASVSGSHGGYDNTALV
SRIMTLRANKAKLGGFTYAAASLENQAKTPEAVNMLGLAPAAVNAKREADLQ
SMIDKEQAKKPLFTKLEAWYSEKVRQAKYNFDESQKPYFELKNVLENGVYFA
AQSEYGLTFKQVTDLPYRDDITVYDVFADGKOLAIFADMYHARGSKRGKAWNSYV
SQSELGTGKPVVANHLPKPPAGQPTLLTWDEVTMFHEPFGHARLHGMVSKYKPYFS
GTSVPRFVPEPSPQVNMWDEPSILKNYKHYQNTGPMQALDKVITAAKSNQGPA
TTEYLGMADQNNWHQISANQVPAAGVMAPEAKALADDGLAYPPVPYKTPYPSHI
MGVGAAYYAIWSEVIDANTQQWFKQHGGLSRANGDRFRQTLISRGSGVDAMELFQ
FAGHAPQITPELLERKGLSAQGGDGTATPEAPQSKP"
2517..3002
/genes="gpo"
/notes="synonym: XCC1413"

```

```
CDS
2517..3002
/gene="gpo"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="glutathione peroxidase-like protein"
/protein_id="AAM40710.1"
/db_xref="GI:21112477"
/translation="MTTSTVQFTDLDPQSLAEYAGKVLVNVVASKGFTPOYAG
LOALQVRRGLVIGPCDFQGHQBPQGAADIRQPCSLDYAVDFPLAAKVEVNGSG
AHLWQWLKHEKRVLSEALKWNTFKELIGRDSVLERYATPKPALLADIALG
A"
complement(3161..3940)
/gene="fpr"
/note="synonym: XCC1414"
complement(3161..3940)
/gene="fpr"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="ferredoxin-NADP reductase"
/protein_id="AAM40711.1"
/db_xref="GI:21112478"
/translation="MSSAPGAETVLEVRHTDVFSTTTTRDAGFRFNGQFVIGLE
TETPLRAYSIAGNWEHEFFSIKVDQGLTSRLQHIQPGDKVLVGKXPTGLLI
SDJHPLRYLIGTGTGLAPWLSIIKDPETVERDFKVLITGHVRFKDLAYDFERE
LPOHEFLGELLHDLKLVYPAVTRAFANQGRLLTBLMDGRMQQTGLPLDPANDRFM
ICGSPQLMADRLTLKDARGFTSPRIPTGPHVVERAFVEX"
complement(4164..6089)
/gene="msba"
/note="synonym: XCC1415"
complement(4164..6089)
/gene="msba"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="AAM40712.1"
/db_xref="GI:21112479"
/translation="MAPNTPSARGPRHTMPSASRPGASAPAPRTGPTLRRERDALEN
LPPFLRHQTSRWLSASSIGLRLVLRALPMTATLYIGKLIIDEALHLVQGPFAFSLT
QALASGRNLRIELALELALAGSLDLGLVGYADALLSELFNNAASVRLMEHAQL
DLEDPEDRDQKLDARRQTMNMLNSQLFGVQDQAITVASLAVGLVYAPMLIVL
LAIALIPAFIEAHFALGYSLNFTWRQLDYLROVGASVETAKVKIINLHFL
ITVRLKDFEQANALARKALWGALLAAGTLGYAAYIAWRTVPGDFSGDL
TFLAGSLRLQLLEGLIGPSQVAGQALYLLDLYSPFNIVPIESRNALPVPEPIR
QGVPEDVGFYPEAEHWTMOHLNPELRAGEVLALVGENGAGKTLVLKLLARLDPDE
GRILLDGHLDRLDYDDVRANGLVIFQDFVYHLSIGENIGVQGDVMDDDTRIKTAA
QRAMATELIDSLPDGYNQLVRRFKNSVDLSGGQWKIAIARAYMRDQMLLDDEFTA
ALDARSFEVFORFELSNDRTAVLISHRFSVRMADRILVLADGRIBESGTHSLMA
OGGRYAELFELQAQYR"
6238..7791
/gene="fumb"
/note="synonym: XCC1416"
6238..7791
/gene="fumb"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="fumarate hydratase"
/protein_id="AAM40713.1"
/db_xref="GI:21112480"
/translation="MEGLPPPAVEPAVTSIKOEDLIQSVDALQVTSYHPVDYIKNL
SAYEREESPAKDAIALIINSRMCASGRPIQDGTGIVTVLEITGMNVDDATMG
VEDVMNVEGRVRYNPDNKLRSVLADPAGKANTKDNPAVSVTKIVPGDHLVIVA
AKGGGSEAKSFAMLNPSDSIVDWLVKTVPTMGACWCPGMLGIGGTAEAKMLLAK
EALMEPIDVLDQAGASNRABELELYKVALGIGAGQGLGLTLDIVDKVDYPT
HAANLVALIPNCAATRAHFTLDGSGPWMLDPPSLSDWPKLTYNPTNRRVNLDTIT
REEVASFKEGSEVLLNGKLLTGRDAAHKRMIDMLNRGETLPVDFTNRTFYVYVGPDPV
```

```
gene
8063..8410
/gene="XCC1417"
8063..8410
/gene="XCC1417"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM40714.1"
/db_xref="GI:21112481"
/translation="MLFSPGNAPQPTAPVPFAEKKPLKPSDTATREDGSPVMERCIA
QCNRSAAQCNADRPSESSCRNAVAPRGKSDATONPAQRANCMQVFDCTQQSDANRC
EPHKLACLRCTTN"
8439..9074
/gene="gst"
/note="synonym: XCC1418"
8439..9074
/gene="gst"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="glutathione S-transferase"
/protein_id="AAM40715.1"
/db_xref="GI:21112482"
/translation="MTTTLXHPSTAALVWHMLIELQVPHVLHPLDFDRREQKSPAY
LALNPAGVPTLVLEGGVMTAEAAAIVLHDLHPQAGLAPPGTTPARAAYTQWMEFFCA
NTLQPAFRMFEYDEPAGAEQVDAQAQOKLBAWAQVDAHLQAHGPYLLIGDTLSA
ADFLITLMKSRNMNPKFTDTWPAHHAQAKPAPRETYRREGLTDWT"
9214..9897
/gene="cynT"
/note="synonym: XCC1419"
9214..9897
/gene="cynT"
Query Match 1.2%; Score 20; DB 1; Length 10029;
Best Local Similarity 100.0%; Fred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 808 ATCATCGACGAGGCGATCCA 827
Db 5840 ATCATCGACGAGGCGATCCA 5821
RESULT 12
LOCUS AE013499 11115 bp DNA linear BCT 07-FEB-2003
DEFINITION Methanosarcina mazei strain Goel, section 281 of 379 of the
complete genome.
ACCESSION AE013499 AE008384
VERSION AE013499.1 GI:20907150
KEYWORDS
SOURCE Methanosarcina mazei Goel
ORGANISM Methanosarcina mazei Goel
Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
Methanosarcinales; Methanosarcinaceae; Methanosarcina.
1 (bases 1 to 11115)
Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A.,
Martinez-Arias, R., Henne, A., Wier, A., Baeumer, S., Jacobi, C.,
Bruggemann, H., Lienard, T., Christmann, A., Boencke, M., Steckel, S.,
Bhattacharya, A., Lykidis, A., Overbeek, R., Klenk, H.-P.,
Gunsalus, R. P., Fritz, H.-J. and Gottschalk, G.
The genome of Methanosarcina mazei: evidence for lateral gene
transfer between bacteria and archaea
J. Mol. Microbiol. Biotechnol. 4 (4), 453-461 (2002)
22120827
12125824
2 (bases 1 to 11115)
Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A.,
Martinez-Arias, R., Henne, A., Wier, A., Baeumer, S., Jacobi, C.,
```

Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,
 Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,
 Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.
 Direct Submission
 Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute
 for Microbiology and Genetics, Grisebachstrasse 8, Goettingen
 37077, Germany

FEATURES

source

Location/Qualifiers

1. .11115

/organism="Methanosarcina mazei Goel"

/mol_type="genomic DNA"

/strain="Goel"

/db_xref="taxon:192952"

196. .753

/gene="MM2548"

196. .753

/gene="MM2548"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAM32244.1"

/db_xref="GI:20907151"

/translations="MMAVSIVFFPIATYDTSILGQNLDISLYNTKIMSTYGMTGGE
 DLTEEVLYQLFSDDYTKPKDKVPDYATVIYNMKIPFADILWFLIVFLGSLPFS
 QNRFLVLAGAILLYAVIIVTYFEIIRUTVQNISFSLITGIPDMNGSTYIFKLGY
 GVCVASICIVIKDNPFRFKIRW"
 complement (911. .1171)

/genes="MM2549"

complement (911. .1171)

/gene="MM2549"

/codon_start=1

/transl_table=11

/product="Conserved hypothetical protein"

/protein_id="AAM32245.1"

/db_xref="GI:20907152"

/translations="MYFLKIRSELDKFKLAKNKKQLEIILAKADEILENPHRYKN
 LKAPMHLARVHDHFLVFLVFSIDESRSVTLDDYDHDKIY"
 complement (1176. .1415)

/genes="MM2550"

complement (1176. .1415)

/gene="MM2550"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAM32246.1"

/db_xref="GI:20907153"

/translations="MVQAIINDERTNRLNLIKAKYGLKSAINKMAEYEBEIL
 EPELKPVEIKLKKIEKQAEIVGVENDLKKRYGL"
 1635. .2264

/gene="MM2551"

1635. .2264

/gene="MM2551"

/codon_start=1

/transl_table=11

/product="Acetyltransferase"

/protein_id="AAM32247.1"

/db_xref="GI:20907154"

/translations="MGNGRYHIDLETWKRKEYCQIYRTAVOQVCSELDVTFPKKH
 VKNELSTMAPIFATVTCANEIEFRFRFLDGEVLVESIDASILDKETELFKVV
 NVPMQDTIEEFVQLATVAENQKEHFTGVENDVIOFSALPWITTHISHDFFGNREK
 AQPTDWKGYHKKGKLMWPFVAVQVHHAFVDGIH:GKLADKLQRYMDEV"
 complement (2343. .2978)

/genes="MM2552"

complement (2343. .2978)

/gene="MM2552"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAM32248.1"

/db_xref="GI:20907155"

/translations="MKTIPDGGSMLESRITIAQSDDSYIERMFEGBILSDVDIARI
 EFESVQTCQFCKDFSCAFYNSVFNCFNCFIMESYWKSGSRILTCRKGDSNFS
 KSHMKETFLADGSFRYAFSSLSMENCISRDCNFTAEFLSLRILKKPLTKAVNFAGAD

gene

CDS

FFKLLKGIDLSDCIIDGITVSDTRNELRGAI VNAQGAIEIAARVLGKIY"

3474. .4313

/gene="MM2553"

3474. .4313

/gene="MM2553"

/codon_start=1

/transl_table=11

/product="putative hydrolase"

/protein_id="AAM32249.1"

/db_xref="GI:20907156"

/translations="MLADSTILKTEYGDIEYTVVEGCTPILLHGAAGGYDQGLWAKK
 VFRGDGYKFTSVSRXYGLNSSI PDNASIESQAAAYKTLNLNLSIDKVI VGVSGGGS
 ATQFANDYPRCSALILISAVSGMPAPQDDPFYVSIHIIQQSDYAYWLLTFRQFA
 VLDMGTFQVYTFENPEQKELAQMDIMHPMSQRYRTVNDKEMIEFTYPTDNLG
 APALI IHA KDDALVSYKHAENAHSKINSQVLVYDTGGHGLMQMNGTRVLVKEFLNE
 STY"

4494. .5147

/gene="MM2554"

4494. .5147

/gene="MM2554"

/EC number="2.3.1.1"

/codon_start=1

/transl_table=11

/product="Acetyltransferase"

/protein_id="AAM32250.1"

/db_xref="GI:20907157"

/translations="MKRRDYFRSEKTIIMNISIRVEEKDFKIVEYTTREAFWDLVKP
 GCDEHLVHKIRKVPAPFVKELDFVACEDTIVGNI IYSRAKLVNEENKEFEVLCMGPL
 TVLPYKQKQIGSLIMNYSIEKARQLAYKGVLI FGNPDYVHREGFNAREYDRTSWG
 DNFDAPMALELYDGLSGISGKFEYDEVEFKIENELEDEFEKQFPYKEKHVDITQLKL"
 complement (5252. .5566)

/genes="MM2555"

complement (5252. .5566)

/gene="MM2555"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAM32251.1"

/db_xref="GI:20907159"

/translations="MITQHSLLFFQRPDKWVDPDGMWYEEYQAAADKMLRWAEPPGE
 GIDLHPDSDRRRRIATSGSLI FSOAPLLEEDRLADRLWDSIALPGTSGIVTPEKK
 TH"

5679. .6968

/gene="MM2556"

5679. .6968

/gene="MM2556"

/codon_start=1

/transl_table=11

/product="ATP-dependent RNA helicase"

/protein_id="AAM32252.1"

/db_xref="GI:20907159"

/translations="NFLIYLYKMNMKKOLSSFFRKIMESLTMMNFKLGLSNAVLKSI
 DEHGFEPTAQEKSIPLILAGEDVIAGATSGSKTSLFASGILONSEKSGKI QALIL
 TPTRELAOVANSLRKFSKYDPLNIASIVGGVINGTQIKELKNAEVVGTPTGRLLDHI
 SENTIKANNKTVLVDADHDFMGKVDVEKIKCEPKQNTQILFSATITKIVRLS
 RKMENPVRVSTESYIDEQKLNQVYQVQDDMKLSLLVYLQNEKSNLGMVFCNFKRN
 TDKAVKRNKSNINAVAHGGLQNTERTRILEKPHSGNIGVLVCTDVAGRGDIDIGVS
 HVIYNDIPRESQYIHRIGRTARAGTEGKAINILSKNDHAFMSVLKNDVDIDKEQAL
 PALKKIVTERPDILRKPVRNMNPSRKGKPRHKNK"

/gene="MM2557"

7709. .9406

/gene="MM2557"

/EC number="1.11.1.6"

/codon_start=1

/transl_table=11

/product="Catalase"

/protein_id="AAM32253.1"

/db_xref="GI:20907160"

/translations="MKGDNCKNPMDKNLGELPSRNQTVGDTPELTTVAGAPVESNQ
 DSMTSGRRGGLPLQDIWFLKLAHEDREVI PERBMDHAKGSGAFGTFTVTHDITSYTKA
 AIPSEIKQKQWMPARFSTVAGERGAADAEDIRGFAUKFYTEBGNWDMVGNNITPVFFF
 RDLKFPDLNHAIKPRTNMRSPTNWDFTWLSLPEALLQLTIINGDRGLPSSFRHMH

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics -----
Center project name: H_MS2026H06

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
Clone CTD-2026H6 is from a release of the human BAC library CTD. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBeloBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-483E17, 2000 bp overlap; the clone sequenced to the right is RP11-493G24, 2000 bp overlap. Actual start of this clone is at base position 51873 of RP11-483E17; actual end is at base position 26168 of RP11-493G24.

Single plasmid region exists between 25828 and 25846. Polymorphisms exist between AC010894, AC018890 and AC104595.

FEATURES	source
repeat_region	1..45334 /location=Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="2" /map="2"
misc_feature	/clone="CTD-2026H6" /clone_lib="CTD" 1162..1491 /rpt_family="Alu"
repeat_region	1509..1988 /note="similar to EST AI282534 (NID:G3920767) qv21b01.x1"
repeat_region	2411..2604 /rpt_family="MIR"
repeat_region	2684..2817 /rpt_family="Alu"
repeat_region	3129..3252 /rpt_family="MIR"
misc_feature	3318..3665 /note="similar to EST BI036921 (NID:G14443447)"
repeat_region	3435..3564 /rpt_family="Achoho"
repeat_region	3641..3794 /rpt_family="CRI"
repeat_region	3795..3919 /rpt_family="MIR"
repeat_region	3935..3967 /rpt_family="L2"

GFSSHAYSLINKNNRVMVRPHFRSQOQIRNLTDQEAIVVVMDSHQDLYEAIEK
GMPKMWYQVTEQAGNKNPFDLTKWKKKDFPLMPVGFELNKNPENYFADV
EQAFANVPGISPSDMLQGLFSGDAHYRLGVNHQIPVNSPOQVKYHSF
HRQOMVNGRSQSLHENSIGNWKDQOYALPEVKTGSTDLRYDFRSDSDYVT
QPGELFRAMPDQOQVLPENTARNMGDSTQIKERHLYNCYQADPKYEGVAKALEID
ISTVDLNLPARTSKANYEANNKHPELNQPTMKKDSGREIDTNGDPPIEPENDEPWL
9919..10908
/gene="WM2558"
9919..10908
/gene="WM2558"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA032254.1"
/db_xref="GI:20907161"
/translation="MEERKTFSTFNTVEVEIILEPTFMFGVKELTKTFVPGHDLQAY
LDPTGKGIYDHLFDPIEAVKWKINORLEWHGKLIQAQCKNDKDLVGLTAL
YIIFLLTIIEFVPPVKSIVGVEPFRWEGYRFGDSTSKETQORVIDSG
NLSFTCDLNRLLAAUSFYTKARLRDNRSGQYEFNVAEVLNKKVLEVFSSKKDK
AKQGLSKLGIYDIEDNFITLFDLRNQDVGHSLSLTINQQQLTDLYEYLEDTEHIF
RELRLTVMKVDGSGYQLKRDSDFKPKNEKILKILDEIFPKIQCKRKTNRSTKN"

BASE COUNT 3585 a 1965 c 2248 g 3317 t

Query Match 1.2%; Score 20; DB 1; Length 11115;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 934 TCCAGGATTGATTCCAC 953
DB 1471 TCCAGGATTGATTCCAC 1490

RESULT 13
AC104595/c
LOCUS 45334 bp DNA linear PRI 03-APR-2002
DEFINITION Homo sapiens BAC clone CTD-2026H6 from 2, complete sequence.
ACCESSION AC104595
VERSION AC104595.6 GI:19909429
KEYWORDS HTG.
SOURCE Homo sapiens (human)

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 45334)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
2 (bases 1 to 45334)
Nguyen, C., Kozlowicz, A., Dignan, G. and Pearman, C.
The sequence of Homo sapiens BAC clone CTD-2026H6
Unpublished (2001)
3 (bases 1 to 45334)
Waterston, R.H.
Direct Submission
Submitted (13-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 45334)
Waterston, R.H.
Direct Submission
Submitted (21-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 45334)
Waterston, R.
Direct Submission
Submitted (03-APR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 3, 2002 this sequence version replaced gi:18855129.

COMMENT
Genome Center

```

repeat_region 4210..4443
/rpt_family="MIR"
repeat_region 4802..5319
/rpt_family="L2"
repeat_region 5586..5890
/rpt_family="MER2_type"
repeat_region 7440..7576
/rpt_family="MIR"
repeat_region 8336..8477
/rpt_family="Alu"
repeat_region 8523..8551
/rpt_family="AT_rich"
misc_feature 8772..9234
/notes="similar to EST AU157565 (NID:gl1019086)"
repeat_region 9105..9337
/rpt_family="MER1_type"
repeat_region 9544..9827
/rpt_family="Alu"
repeat_region 10365..10445
/rpt_family="MIR"
repeat_region 11746..11934
/rpt_family="MIR"
repeat_region 12427..12644
/rpt_family="MIR"
repeat_region 12985..13111
/rpt_family="L2"
repeat_region 13145..13434
/rpt_family="Alu"
repeat_region 14651..15745
/rpt_family="ERVk"
misc_feature 15993..16425
/notes="similar to EST AW792849 (NID:g7844771)"
repeat_region 17363..17394
/rpt_family="AT_rich"
repeat_region 17640..17661
/rpt_family="AT_rich"
repeat_region 18305..18417
/rpt_family="L1"
repeat_region 18449..18493
/rpt_family="(CA)n"
repeat_region 18516..18872
/rpt_family="L1"
repeat_region 18891..19198
/rpt_family="Alu"
misc_feature 19072..19695
/notes="similar to EST AL598697 (NID:gl5161388)"
misc_feature 19146..19334
/notes="similar to EST BI048567 (NID:gl14455189)"
repeat_region 19488..19740
/rpt_family="Alu"
repeat_region 19764..19818
/rpt_family="L1"
repeat_region 19819..19857
/rpt_family="(TCCC)n"
repeat_region 19858..20150
/rpt_family="Alu"
misc_feature 20371..20870
/notes="similar to EST AA610841 (NID:g2459269) hp98e08.s1"
repeat_region 20426..20533
/rpt_family="Alu"
repeat_region 20534..20555
/rpt_family="(CAAAA)n"
repeat_region 20556..20708
/rpt_family="Alu"
repeat_region 21424..21590
/rpt_family="MIR"
misc_feature 21447..21853
/notes="similar to EST BF897239 (NID:gl2288698)"
repeat_region 21618..21669
/rpt_family="GA-rich"
misc_feature 22049..22395
/notes="similar to EST AW903959 (NID:g8069164)"
repeat_region 22595..22731

```

```

/rpt_family="MIR"
22734..23027
/rpt_family="Alu"
23091..23179
/rpt_family="MER1_type"
23521..23571
/rpt_family="L2"
24448..24529
/rpt_family="MIR"
25680..26697
/notes="CpG island (%GC=76.0, o/e=0.90, #CpGs=136)"
26484..26580
/rpt_family="GC_rich"
27406..28038
/rpt_family="ERV1"
28422..28557
repeat_region

Query Match 1.2%; Score 20; DB 9; Length 45334;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 CCCTTGTTGCAGCCAGTTG 50
|||||
DB 42972 CCCTTGTTGCAGCCAGTTG 42953

RESULT 14
AP004005/c
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
BAC clone:OJ1188_F05, complete sequence.
ACCESSION AP004005
VERSION AP004005.2 GI:28208278
SOURCE HTG.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
clone:OJ1188_F05
JOURNAL Published Only in Database (2001)
REFERENCE 2 (bases 1 to 82843)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-299-38-7441, Fax:81-298-38-7468)
On Feb 4, 2003 this sequence version replaced gi:15076813.
The orientation of the sequence is from -21M13 to M13rev of the BAC
clone. The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
FEATURES
Location/Qualifiers
1..82843
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="8"
/clone="OJ1188_F05"
BASE COUNT 23050 a 17555 c 17060 g 25178 t
ORIGIN

Query Match 1.2%; Score 20; DB 8; Length 82843;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 893 TCTACAACTACATGAAGTTT 912
|||||

```

Db 29092 TCTACAACTACATGAGTTT 29073

Query Match 1.2%; Score 20; DB 8; Length 96161;
Best Local Similarity 100.0%; Pred.No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15

AB062091 96161 bp DNA linear PLN 26-MAR-2003
LOCUS Arabidopsis thaliana DNA, chromosome 1 centromere region, HAC
DEFINITION Clone:F9A12.

ACCESSION AB062091

VERSION AB062091.1

KEYWORDS GI:18148664

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 Hosouchi, T., Kume, K., Tsuruoka, H. and Kotani, H.

Physical map-based sizes of the centromeric regions of Arabidopsis

thaliana chromosomes 1, 2, and 3

JOURNAL DNA Res. 9 (4), 117-121 (2002)

MEDLINE 22225143

PUBMED 12240833

REFERENCE 2 (bases 1 to 96161)

AUTHORS Kotani, H. and Kume, K.

TITLE Direct Submission

Submitted (21-MAY-2001) Hirokazu Kotani, Kazusa DNA Research

Institute, Lab. Chromosome Research II; 1532-3 Yana, Kisarazu,

Chiba 292-0812, Japan (E-mail: kotani@kazusa.or.jp,

Tel: 81-438-52-3905, Fax: 81-438-52-3924)

Location/Qualifiers

FEATURES

source

1..96161

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia"

/db_xref="taxon:3702"

/chromosome="1"

/map="centromere"

/clone="F9A12"

1..4459

misc_feature

/note="similarity to fragments of Tat retrotransposons"

LTR

repeat_region

4460..4962

/note="similarity to 550 bp repeat"

LTR

misc_feature

23573..24405

/note="similarity to fragments of retrotransposons"

LTR

repeat_region

26928..27741

/note="similarity to 550 bp repeat"

LTR

repeat_region

36974..37268

/note="similarity to 550 bp repeat"

LTR

misc_feature

64191..71814

/note="similarity to fragments of Tat retrotransposons"

LTR

repeat_region

71815..72541

/note="similarity to fragments of En/Spm transposon"

misc_feature

75204..79725

/note="similarity to 550 bp repeat"

repeat_region

79726..83303

/note="similarity to fragments of En/Spm transposon"

misc_feature

83304..90269

/note="similarity to 550 bp repeat"

LTR

repeat_region

90270..91072

/note="similarity to fragments of Tat retrotransposons"

misc_feature

91073..95051

/note="similarity to 550 bp repeat"

BASE COUNT 31185 a 23079 c 13560 g 28337 t

ORIGIN

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus p2n model

Run on: November 21, 2003, 20:29:44 ; Search time 475 Seconds
(without alignments)
3818.639 Million cell updates/sec

Title: US-09-712-338-2

Perfect score: 555

Sequence: 1 MRGYEFLSVLPLAASWALP.....HTQSSVPLPTATSSVSGMA 555

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4333170

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09712338/runat_17112003_170559_11758/app.query.fasta_1.711
-DB=published Applications NA -QFMT=fastap -SUFFIX=oli.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09712338@cgn 1.1.107 @runat_17112003_170559_11758
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.6	586	12	US-10-027-632-279095, Sequence 279095,

2	1.6	586	12	US-10-027-632-279096	Sequence 279096,
3	1.6	586	13	US-10-027-632-279095	Sequence 279095,
4	1.6	586	13	US-10-027-632-279096	Sequence 279096,
5	1.6	1575	14	US-10-156-761-2224	Sequence 2224, Ap
6	1.6	2581	12	US-10-027-632-249356	Sequence 249356,
7	1.6	2581	12	US-10-027-632-249357	Sequence 249357,
8	1.6	2581	12	US-10-027-632-249358	Sequence 249358,
9	1.6	2581	12	US-10-027-632-249359	Sequence 249359,
10	1.6	2581	12	US-10-027-632-249360	Sequence 249360,
11	1.6	2581	13	US-10-027-632-249356	Sequence 249356,
12	1.6	2581	13	US-10-027-632-249357	Sequence 249357,
13	1.6	2581	13	US-10-027-632-249358	Sequence 249358,
14	1.6	2581	13	US-10-027-632-249359	Sequence 249359,
15	1.6	2581	13	US-10-027-632-249360	Sequence 249360,
16	1.6	175590	11	US-09-911-077A-13	Sequence 13, Appl
17	1.6	9025608	14	US-10-156-761-1	Sequence 1, Appl
18	1.4	173	10	US-09-878-574-11661	Sequence 11661, A
19	1.4	228	9	US-09-864-761-30286	Sequence 30286, A
20	1.4	370	10	US-09-998-598-2056	Sequence 2056, Ap
21	1.4	370	10	US-09-998-598-2497	Sequence 2497, Ap
22	1.4	372	9	US-09-764-869-170	Sequence 170, App
23	1.4	372	10	US-09-860-670-38	Sequence 38, Appl
24	1.4	372	14	US-10-091-504-170	Sequence 170, App
25	1.4	397	12	US-10-027-632-60707	Sequence 60707, A
26	1.4	397	12	US-10-027-632-300964	Sequence 300964, A
27	1.4	397	13	US-10-027-632-60707	Sequence 60707, A
28	1.4	397	13	US-10-027-632-300964	Sequence 300964, A
29	1.4	410	11	US-09-918-995-5613	Sequence 5613, Ap
30	1.4	423	12	US-10-027-632-182906	Sequence 182906,
31	1.4	423	12	US-10-027-632-182907	Sequence 182907,
32	1.4	423	13	US-10-027-632-182906	Sequence 182906,
33	1.4	423	13	US-10-027-632-182907	Sequence 182907,
34	1.4	472	11	US-09-918-995-9285	Sequence 9285, Ap
35	1.4	481	10	US-09-878-574-4508	Sequence 4508, Ap
36	1.4	483	10	US-09-833-381-680	Sequence 680, App
37	1.4	501	10	US-09-796-692-7738	Sequence 7738, Ap
38	1.4	501	14	US-10-040-862-7738	Sequence 7738, Ap
39	1.4	503	11	US-09-918-995-27406	Sequence 27406, A
40	1.4	508	10	US-09-796-692-7829	Sequence 7829, Ap
41	1.4	508	14	US-10-040-862-7829	Sequence 7829, Ap
42	1.4	512	10	US-09-867-701-2137	Sequence 2137, Ap
43	1.4	536	10	US-09-796-692-7558	Sequence 7558, Ap
44	1.4	536	14	US-10-040-862-7558	Sequence 7558, Ap
45	1.4	557	10	US-09-920-455-120	Sequence 120, App

ALIGNMENTS

RESULT 1

US-10-027-632-279095
; Sequence 279095, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 279095
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-279095

Alignment Scores:
Pred. No.: 10.4 Length: 586
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 12 Gaps: 0

US-09-712-338-2 (1-555) x US-10-027-632-279095 (1-586)

QY 416 AspLeuGluGluLeuAlaLeuPro 424
Db 98 GATCTTGAAGAGATATTAGCACTTCCA 124

RESULT 2

US-10-027-632-279096
; Sequence 279096, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 279096
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-279096

Alignment Scores:
Pred. No.: 10.4 Length: 586
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 12 Gaps: 0

US-09-712-338-2 (1-555) x US-10-027-632-279096 (1-586)

QY 416 AspLeuGluGluLeuAlaLeuPro 424
Db 98 GATCTTGAAGAGATATTAGCACTTCCA 124

RESULT 3

US-10-027-632-279095
; Sequence 279095, Application US/10027632
; GENERAL INFORMATION:

APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 279095
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-279095

Alignment Scores:

Pred. No.: 10.4 Length: 586
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 13 Gaps: 0

US-09-712-338-2 (1-555) x US-10-027-632-279095 (1-586)

QY 416 AspLeuGluGluLeuAlaLeuPro 424
Db 98 GATCTTGAAGAGATATTAGCACTTCCA 124

RESULT 4

US-10-027-632-279096
; Sequence 279096, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 279096
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-279096

Alignment Scores:
Pred. No.: 10.4 Length: 586
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 13 Gaps: 0

US-09-712-338-2 (1-555) x US-10-027-632-279096 (1-586)

QY 416 AspleuGlulLeuAlaLeuPro 424

Db 98 GATCTTGAAGAGATATTAGCACTTCCA 124

RESULT 5

US-10-156-761-2224
; Sequence 2224, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 13109
; SEQ ID NO 2224
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1575)
US-10-156-761-2224

Alignment Scores:
Pred. No.: 26.1 Length: 1575
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 14 Gaps: 0

US-09-712-338-2 (1-555) x US-10-156-761-2224 (1-1575)

QY 14 AlaAlaSerTrpAlaLeuProGlySer 22

Db 224 GCCGCGTCTGGCGCTCCCGCATCC 250

RESULT 6

US-10-027-632-249356
; Sequence 249356, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249356
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249356

Alignment Scores:

Pred. No.: 41.3 Length: 2581
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 12 Gaps: 0

US-09-712-338-2 (1-555) x US-10-027-632-249356 (1-2581)

QY 207 PheLeuSerGlyLeuProSerLeuAsp 215

Db 2001 TTCTCTCTGGTCTTCCCTCTGGAC 2027

RESULT 7

US-10-027-632-249357
; Sequence 249357, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249357
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249357

Alignment Scores:

Pred. No.: 41.3 Length: 2581
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 12 Gaps: 0

US-09-712-338-2 (1-555) x US-10-027-632-249357 (1-2581)

QY 207 PheLeuSerGlyLeuProSerLeuAsp 215
Db 2001 TTCCTCTCTGGTCTTCCCTCTCTGGAC 2027

RESULT 8

US-10-027-632-249358
; Sequence 249358, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249358
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249358

Alignment Scores:
Pred. No.: 41.3 Length: 2581
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
Gaps: 12

US-09-712-338-2 (1-555) x US-10-027-632-249358 (1-2581)

QY 207 PheLeuSerGlyLeuProSerLeuAsp 215
Db 2001 TTCCTCTCTGGTCTTCCCTCTCTGGAC 2027

RESULT 9

US-10-027-632-249359
; Sequence 249359, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249359
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249359

Alignment Scores:
Pred. No.: 41.3 Length: 2581
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
Gaps: 12

US-09-712-338-2 (1-555) x US-10-027-632-249359 (1-2581)

QY 207 PheLeuSerGlyLeuProSerLeuAsp 215
Db 2001 TTCCTCTCTGGTCTTCCCTCTCTGGAC 2027

RESULT 10

US-10-027-632-249360
; Sequence 249360, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249360
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249360

Alignment Scores:
Pred. No.: 41.3 Length: 2581
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
Gaps: 12

US-09-712-338-2 (1-555) x US-10-027-632-249360 (1-2581)

QY 207 PheLeuSerGlyLeuProSerLeuAsp 215
Db 2001 TTCCTCTCTGGTCTTCCCTCTCTGGAC 2027

```
Db      2001 TTCCCTCTGGTCTTCCCTCTCTGGAC 2027
RESULT 11
US-10-027-632-249356
; Sequence 249356, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249356
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249356
Alignment Scores:
Pred. No.:      41.3      Length:      2581
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     1.62%      Indels:      0
DB:              13      Gaps:        0

US-09-712-338-2 (1-555) x US-10-027-632-249356 (1-2581)
QY      207 PheLeuSerGlyLeuProSerLeuAsp 215
      |||||
Db      2001 TTCCCTCTGGTCTTCCCTCTCTGGAC 2027
RESULT 12
US-10-027-632-249357
; Sequence 249357, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249357
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249357
Alignment Scores:
Pred. No.:      41.3      Length:      2581
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     1.62%      Indels:      0
DB:              13      Gaps:        0

US-09-712-338-2 (1-555) x US-10-027-632-249356 (1-2581)
QY      207 PheLeuSerGlyLeuProSerLeuAsp 215
      |||||
Db      2001 TTCCCTCTGGTCTTCCCTCTCTGGAC 2027
RESULT 13
US-10-027-632-249358
; Sequence 249358, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249358
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249358
Alignment Scores:
Pred. No.:      41.3      Length:      2581
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     1.62%      Indels:      0
DB:              13      Gaps:        0

US-09-712-338-2 (1-555) x US-10-027-632-249358 (1-2581)
QY      207 PheLeuSerGlyLeuProSerLeuAsp 215
      |||||
Db      2001 TTCCCTCTGGTCTTCCCTCTCTGGAC 2027
RESULT 14
US-10-027-632-249359
; Sequence 249359, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249357
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249357
Alignment Scores:
Pred. No.:      41.3      Length:      2581
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     1.62%      Indels:      0
DB:              13      Gaps:        0

US-09-712-338-2 (1-555) x US-10-027-632-249357 (1-2581)
QY      207 PheLeuSerGlyLeuProSerLeuAsp 215
      |||||
Db      2001 TTCCCTCTGGTCTTCCCTCTCTGGAC 2027
RESULT 13
US-10-027-632-249358
; Sequence 249358, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249358
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249358
Alignment Scores:
Pred. No.:      41.3      Length:      2581
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     1.62%      Indels:      0
DB:              13      Gaps:        0

US-09-712-338-2 (1-555) x US-10-027-632-249358 (1-2581)
QY      207 PheLeuSerGlyLeuProSerLeuAsp 215
      |||||
Db      2001 TTCCCTCTGGTCTTCCCTCTCTGGAC 2027
RESULT 14
US-10-027-632-249359
; Sequence 249359, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

;; TITLE OF INVENTION: Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 249359
;; LENGTH: 2581
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-249359

Alignment Scores:
Pred. No.: 41.3 Length: 2581
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 13 Gaps: 0

US-09-712-338-2 (1-555) x US-10-027-632-249359 (1-2581)

QY 207 PheLeuSerGlyLeuProSerLeuAasp 215
|||
DB 2001 TTCCCTCTCTGGTCTTCCCTCTCTGGAC 2027

RESULT 15

US-10-027-632-249360
; Sequence 249360, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249360
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-249360

Alignment Scores:

Pred. No.: 41.3 Length: 2581
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 13 Gaps: 0

US-09-712-338-2 (1-555) x US-10-027-632-249360 (1-2581)

QY 207 PheLeuSerGlyLeuProSerLeuAasp 215
|||
DB 2001 TTCCCTCTCTGGTCTTCCCTCTCTGGAC 2027

Search completed: November 21, 2003, 22:51:16
Job time : 492 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 20:21:34 ; Search time 101 Seconds
(without alignments)
2425.422 Million cell updates/sec

Title: US-09-712-338-2

Perfect score: 555

Sequence: 1 MRGYEFLSVPLVAASWALP.....HTQSSVPLPTATSMSSVGMA 555

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1135299

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09712338/runat_17112003_170559_11735/app_query.fasta_1.711
-DB=Issued Patents NA -QPM=fastap -SUFFIX=oli.rni -MINMATCH=0.1 -LOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCLALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CGN 1 1 103 @runat 17112003 170559 11735 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	554	99.8	1662	3	US-08-343-714-1
2	9	1.6	239	4	US-09-313-294A-6277
3	9	1.6	55216	4	US-09-716-865-23
4	9	1.6	55216	4	US-09-716-865-23
5	8	1.4	236	4	US-09-313-294A-6788
6	8	1.4	471	4	US-09-252-991A-8611
7	8	1.4	497	4	US-09-439-554-1
8	8	1.4	630	4	US-09-252-991A-10808
9	8	1.4	852	4	US-09-107-532A-1764
10	8	1.4	882	4	US-09-252-991A-10854
11	8	1.4	930	4	US-09-252-991A-10471
12	8	1.4	1069	4	US-09-482-273-22

13	8	1.4	1118	4	US-09-439-554-19	Sequence 19, Appl
14	8	1.4	1154	4	US-09-482-273-86	Sequence 86, Appl
15	8	1.4	1197	4	US-09-482-273-87	Sequence 87, Appl
16	8	1.4	1395	4	US-08-976-259-99	Sequence 99, Appl
17	8	1.4	1761	4	US-09-252-991A-10827	Sequence 10827, A
18	8	1.4	1814	4	US-09-702-705-319	Sequence 319, App
19	8	1.4	1814	4	US-09-736-457-319	Sequence 319, App
20	8	1.4	2002	1	US-08-309-341-3	Sequence 3, Appl
21	8	1.4	2002	1	US-08-608-267-3	Sequence 3, Appl
22	8	1.4	2002	1	US-08-608-452-3	Sequence 3, Appl
23	8	1.4	2002	1	US-08-608-224-3	Sequence 3, Appl
24	8	1.4	2002	2	US-08-967-149-3	Sequence 3, Appl
25	8	1.4	2068	1	US-08-309-341-1	Sequence 1, Appl
26	8	1.4	2068	1	US-08-608-267-1	Sequence 1, Appl
27	8	1.4	2068	1	US-08-608-452-1	Sequence 1, Appl
28	8	1.4	2068	1	US-08-608-224-1	Sequence 1, Appl
29	8	1.4	2068	2	US-08-967-149-1	Sequence 1, Appl
30	8	1.4	2084	4	US-08-646-242-7	Sequence 7, Appl
31	8	1.4	2503	1	US-09-640-305-3	Sequence 3, Appl
32	8	1.4	2503	1	US-08-360-673-3	Sequence 3, Appl
33	8	1.4	2632	2	US-08-899-324-32	Sequence 32, Appl
34	8	1.4	2632	3	US-08-329-892B-32	Sequence 32, Appl
35	8	1.4	2668	4	US-09-370-838-156	Sequence 156, App
36	8	1.4	2859	2	US-08-506-340A-2	Sequence 2, Appl
37	8	1.4	3244	2	US-08-321-670-1	Sequence 1, Appl
38	8	1.4	17949	3	US-09-087-465-3	Sequence 3, Appl
39	8	1.4	26173	4	US-09-453-702B-69	Sequence 69, Appl
40	8	1.4	26700	1	US-08-472-217-1	Sequence 1, Appl
41	8	1.4	26700	2	US-08-488-199-5	Sequence 5, Appl
42	8	1.4	26700	3	US-08-760-534A-1	Sequence 1, Appl
43	8	1.4	26700	4	US-09-336-757-1	Sequence 1, Appl
44	8	1.4	36181	4	US-08-311-731A-120	Sequence 120, App
45	7	1.3	26	3	US-09-117-217-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-08-943-714-1
; Sequence 1, Application US/08943714
; Patent No. 6187578

GENERAL INFORMATION:
APPLICANT: Blinkovsky, Alexander

APPLICANT: Berk, Randy

APPLICANT: Rey, Michael

APPLICANT: Golightly, Elizabeth

APPLICANT: Klotz, Alan

APPLICANT: Mathisen, Thomas Erik

APPLICANT: Danbmann, Claus

TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids

TITLE OF INVENTION: Encoding Same

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 61875780 No. 6187578disk of No. 6187578th America, Inc.

STREET: 405 Lexington Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/943,714

FILING DATE: 03-OCT-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4990.200-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1662 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 1...1662
 OTHER INFORMATION:
 US-08-943-714-1

Alignment Scores:
 Pred. No.: 0 Length: 1662
 Score: 554.00 Matches: 554
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.82% Indels: 0
 DB: 3 Gaps: 0

US-09-712-338-2 (1-555) x US-08-943-714-1 (1-1662)

QY 1 MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuPro 20
 Db 1 ATCGTGGCTACGAATTTCTCTCAGTGTACCTTGTGGTGCACCGAGTTGGGCCCTTCCA 60
 QY 21 GlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGlyValLys 40
 Db 61 GGAAGTACACCGCGCTCGTGTAGAGACAGCTACCCAGAAACCCACCGGGGTCAAG 120
 QY 41 ThrLeuThrThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal 60
 Db 121 ACTCTTACACCGCAACAATGTACCATCCGTACAGAAACCCCGGGGAGGGGGTCC 180
 QY 61 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer 80
 Db 181 TGGAGACTACCGGGGTGCAATCTCTCTGGATATGTCGACACCTCTCCCGAGTCC 240
 QY 81 HisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 100
 Db 241 CATACCTTCTCTGTTCTTTCGAAGCCAGACATAAACCCAGAAATGCAACCTATCACATTG 300
 QY 101 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluGlyPro 120
 Db 301 TGGTTGAATCGTGGCCCTGGAAGCGATCTTTGATCGGTCTCTTCGAAGAGTTGGGCCCT 360
 QY 121 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 140
 Db 361 TGGCATGTCAATTCGACTTTTGATGACTACATCAACCTCCTCTGTAACGAGGTCTCC 420
 QY 141 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 160
 Db 421 AATTTACTATTCCTGTCCAGCCATTGGGAGTGGGCTTTTTCATATAGTGATACGGTTGAT 480
 QY 161 GlySerIleAsnProValThrGlyValGluAsnSerSerPheAlaGlyValGlnGly 180
 Db 481 GGGTCCATTAACCTGTAACCTGGGGTCTCGAAATTCGACTTTGAGAGGTTCAGGGC 540
 QY 181 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAla 200
 Db 541 CGGTACCAACCAATGATGACCTCTGATGATACCTACCAATCTTTCGCGAGAGCGCGCT 600
 QY 201 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 220
 Db 601 TGGGAGATCTCTGCAAGGATTCCTTAGTGACTACTAGCTTGGACTCTTAGGGTGCAGTCT 660
 QY 221 LysAspPheSerLeuThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn 240
 Db 661 AAGGACTTCAGTCTATGACCGAGAGAGCTATCGAGGGGCACTATGTGCTCTGCAATCTTCAAT 720

QY 241 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 260
 Db 721 CATTTTTCAGAGCAGATTCGCAACGGTAGTGTATTAATGGTGTTCAGCTTAAT 780
 QY 261 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleAlaProTyrTyr 280
 Db 781 TTCAACTCTCTCTGGAAATTTAACGGCATCATCGAGAGCGCATCCAGGCCCTTACTAC 840
 QY 281 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 300
 Db 841 CCTGTAATTCGCTGTGAACAATACCTACGTTATCAAGGCTGTCAACGAGACCGCTACAAC 900
 QY 301 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 320
 Db 901 TACATGAAGTTTGCACCAACCAATGCAATGTTGCCAGGATTTGATTTCCACCTCAAA 960
 QY 321 GlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys 340
 Db 961 CAGACAAACCGCACCGCATTAGCTGACTACGCCCTCTGCGCGAGAGCCACCAACATGTC 1020
 QY 341 ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 360
 Db 1021 AGGACAAATGTTGAGGGGCCATACCTACGCCCTTTGCTGTGCTGTGTATGATATCGG 1080
 QY 361 HisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSer 380
 Db 1081 CATCCATATGATGACCCGACTCCGCCAAGTTATTACAAATAATTTCTGGCAAGACTCT 1140
 QY 381 ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 400
 Db 1141 GTCATGAGCGCTATCGCGGTCAACATCACTACACCCAGTCCATATTAATGACGCTACTAC 1200
 QY 401 AlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle 420
 Db 1201 GCTTTCCAGCAACAGCGACTTTGTCTGGCCCACTTCATCGAAGACCTCGAGGAGATC 1260
 QY 421 LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp 440
 Db 1261 CTTGTCTCTCCCGTGGTGTCTCCCTCATCTATGGCGACCGCATTACATCTGCAACTGG 1320
 QY 441 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 460
 Db 1321 TTGGCGGTTCAGCGCGTTTCCCTCGCTGCGAATCTCTCCCAAGCCGCCAGTTCGGAAGC 1380
 QY 461 AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 480
 Db 1381 GCAGGGTACACCGCCCTGAAAGTCAACGGCGTCGAGTATGGGAAACTCCGAGTATGGT 1440
 QY 481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 500
 Db 1441 AATTTCTCTCTCTCGCTGCTATGAGCGAGGCGATGAATGCCATCTACTACTACCGCCATC 1500
 QY 501 AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys 520
 Db 1501 GCCTCCCTGCAATTTGTTAACCGGACTATCTCGTTGGGATATCGCAGAGGGGCCAGAAG 1560
 QY 521 LysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 540
 Db 1561 AAGATCTGGCCAGCTACAAGACGAATGGAACGGCTACAGCTACGCTACGCTACACAGTCTGCC 1620
 QY 541 ValProLeuProThrAlaThrSerMetSerSerValGlyMet 554
 Db 1621 GTGCGCTGCTCCTACGGCTACGCAATGTCAGTGTTCGATG 1662

RESULT 2

US-09-313-294A-6277
 ; Sequence 6277, Application US/09313294A
 ; Patent No. 6476212
 ; GENERAL INFORMATION:
 ; APPLICANT: Ialguidi, Raghunath V.
 ; APPLICANT: Ito, Laura Y.
 ; APPLICANT: Sherman, Bradley K.


```
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313/294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 6277
; LENGTH: 299
; TYPE: DNA
; ORGANISM: Zea mays
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700351528H1
US-09-313-294A-6277

Alignment Scores:
Pred. No.: 2.15 Length: 299
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-313-294A-6277 (1-299)

QY 540 ServValProLeuProThraAlaThrSer 548
Db 48 TCTGTGCCACTGGCCCACTGCCACCTCA 74

RESULT 3
US-09-716-865-23
; Sequence 23, Application US/09716865
; Patent No. 6548292
; GENERAL INFORMATION:
; APPLICANT: Bramucci, Micheal G
; APPLICANT: Nagarajan, Vasantha
; APPLICANT: Chen, Mario W.
; TITLE OF INVENTION: Bacterial plasmid Having Genes Encoding Enzymes for the
; FILE REFERENCE: BC1016 US NA
; CURRENT APPLICATION NUMBER: US/09/716,865
; CURRENT FILING DATE: 2000-11-26
; PRIOR APPLICATION NUMBER: 60/167,062
; PRIOR FILING DATE: 1999-11-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
; LENGTH: 55216
; TYPE: DNA
; ORGANISM: Pseudomonas CT14
US-09-716-865-23

Alignment Scores:
Pred. No.: 319 Length: 55216
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-716-865-23 (1-55216)

QY 88 GluAlaArgHisAsnProGluThraAla 96
Db 13516 GAAGCGGACACAAACCGGAGCGGCC 13542

RESULT 4
US-09-716-865-23/c
; Sequence 23, Application US/09716865
; Patent No. 6548292
; GENERAL INFORMATION:
; APPLICANT: Bramucci, Micheal G
; APPLICANT: Nagarajan, Vasantha
```

```
; APPLICANT: Chen, Mario W.
; TITLE OF INVENTION: Bacterial Plasmid Having Genes Encoding Enzymes for the
; TITLE OF INVENTION: Degradation of Aromatic Compounds
; FILE REFERENCE: BC1016 US NA
; CURRENT APPLICATION NUMBER: US/09/716,865
; CURRENT FILING DATE: 2000-11-26
; PRIOR APPLICATION NUMBER: 60/167,062
; PRIOR FILING DATE: 1999-11-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
; LENGTH: 55216
; TYPE: DNA
; ORGANISM: Pseudomonas CT14
US-09-716-865-23

Alignment Scores:
Pred. No.: 319 Length: 55216
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-716-865-23 (1-55216)

QY 88 GluAlaArgHisAsnProGluThraAla 96
Db 49740 GAAGCGGACACAAACCGGAGCGGCC 49714

RESULT 5
US-09-313-294A-6788
; Sequence 6788, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 6788
; LENGTH: 296
; TYPE: DNA
; ORGANISM: Zea mays
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700352406H1
; NAME/KEY: unsure
; LOCATION: 54
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-6788

Alignment Scores:
Pred. No.: 23.2 Length: 296
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-313-294A-6788 (1-296)

QY 207 PheLeuSerGlyLeuProSerLeu 214
Db 266 TTCCTCTCCGGCCTCCCTCCCTC 289

RESULT 6
US-09-252-991A-8611
; Sequence 8611, Application US/09252991A
```

US-09-712-338-2 (1-555) x US-09-252-991A-8611 (1-471)

QY 38 GlyValLysThrLeuThrAla 45
Db 305 GGCGTGAAGACCTTGACGACGCA 328

RESULT 7

US-09-439-554-1
; Sequence 1, Application US/09439554
; Patent No. 6479733
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Jan Antoni
; APPLICANT: Odell, Joan T.
; APPLICANT: Sakai, Hajime
; APPLICANT: Thorpe, Catherine J.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Famodu, Oluolayo O.
; TITLE OF INVENTION: STEROL METABOLISM ENZYMES
; FILE REFERENCE: BB1114 US NA
; CURRENT APPLICATION NUMBER: US/09/439,554
; CURRENT FILING DATE: 1999-11-12
; EARLIER APPLICATION NUMBER: 60/108,351
; EARLIER FILING DATE: 1998-NO. 6479733ember-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Zea mays
US-09-439-554-1

Alignment Scores:
Pred. No.: 38.2 Length: 497
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-439-554-1 (1-497)

QY 207 PheLeuSerGlyLeuProSerIleu 214
Db 175 TTCCTCTCGGCGCTCCCTCCCTC 198

RESULT 8

US-09-712-338-2 (1-555) x US-09-252-991A-10808 (1-630)

QY 18 AlaLeuProGlySerThrProAla 25
Db 418 GCCTACCAAGATCATCGCGGCT 395

RESULT 9

US-09-107-532A-1764
; Sequence 1764, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lyvnn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1764:
; SEQUENCE CHARACTERISTICS:

US-09-252-991A-10808/c
; Sequence 10808, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10808
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10808

Alignment Scores:
Pred. No.: 47.9 Length: 630
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-252-991A-10808 (1-630)

QY 18 AlaLeuProGlySerThrProAla 25
Db 418 GCCTACCAAGATCATCGCGGCT 395

RESULT 9

US-09-107-532A-1764
; Sequence 1764, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lyvnn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1764:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 852 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecium
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (3) LOCATION 1...852
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 1764:
 US-09-107-532A-1764

Alignment Scores:
 Pred. No.: 64 Length: 852
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-107-532A-1764 (1-852)

Qy 379 AspSerValMetAspAlaIleGly 386
 Db 559 GATAGTGTGATGATGCAATCGGT 582

RESULT 10

US-09-252-991A-10854/c
 ; Sequence 10854, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 10854
 ; LENGTH: 882
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-10854

Alignment Scores:
 Pred. No.: 66.1 Length: 882
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-252-991A-10854 (1-882)

Qy 21 GlySerThrProAlaSerValGly 28
 Db 383 GGCAGCAGCCGCCAGCGTCGGA 360

RESULT 11

US-09-252-991A-10471/c
 ; Sequence 10471, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 10471
 ; LENGTH: 930
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-10471

; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 10471
 ; LENGTH: 930
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-10471

Alignment Scores:
 Pred. No.: 69.6 Length: 930
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-252-991A-10471 (1-930)

Qy 420 IleLeuAlaLeuProValArgVal 427
 Db 691 ATACTCGCCCTACCGGTGAGATC 668

RESULT 12

US-09-482-273-22
 ; Sequence 22, Application US/09482273
 ; Patent No. 6534631
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: 71 Human Secreted Proteins
 ; FILE REFERENCE: P2030P1
 ; CURRENT APPLICATION NUMBER: US/09/482,273
 ; CURRENT FILING DATE: 2000-01-13
 ; EARLIER APPLICATION NUMBER: PCT/US99/15849
 ; EARLIER FILING DATE: 1999-07-14
 ; EARLIER APPLICATION NUMBER: 60/092,921
 ; EARLIER FILING DATE: 1998-07-15
 ; EARLIER APPLICATION NUMBER: 60/092,922
 ; EARLIER FILING DATE: 1998-07-15
 ; EARLIER APPLICATION NUMBER: 60/092,956
 ; EARLIER FILING DATE: 1998-07-15
 ; NUMBER OF SEQ ID NOS: 267
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 22
 ; LENGTH: 1069
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (508)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-09-482-273-22

Alignment Scores:
 Pred. No.: 79.5 Length: 1069
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-482-273-22 (1-1069)

Qy 423 LeuProValArgValSerLeulle 430
 Db 278 CTCCTGTCGCGGTGCTCTGATC 301

RESULT 13

US-09-439-554-19
; Sequence 19, Application US/09439554
; Patent No. 6479733
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Jan Antoni
; APPLICANT: Odell, Joan T.
; APPLICANT: Sakai, Hajime
; APPLICANT: Thorpe, Catherine J.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: STEROL METABOLISM ENZYMES
; FILE REFERENCE: BB1114 US NA
; CURRENT APPLICATION NUMBER: US/09/439,554
; CURRENT FILING DATE: 1999-11-12
; EARLIER APPLICATION NUMBER: 60/108,351
; EARLIER FILING DATE: 1998-No. 6479733ember-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 1118
; TYPE: DNA
; ORGANISM: Zea mays
US-09-439-554-19

Alignment Scores:
Pred. No.: 83 Length: 1118
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-439-554-19 (1-1118)

QY 207 PheLeuSerClyLeuProSerLeu 214
DB 183 TTCCTCTCGGCTCCCTCCCTCCCTC 206

RESULT 14
US-09-482-273-86
; Sequence 86, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 86
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-482-273-86

Alignment Scores:
Pred. No.: 85.6 Length: 1154
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-482-273-86 (1-1154)

QY 423 LeuProValArgValSerLeuIle 430
DB 354 CTCCTGTCCGTGTGTCCTTGATC 377

RESULT 15
US-09-482-273-87
; Sequence 87, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 87
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (573)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1177)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1185)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-482-273-87

Alignment Scores:
Pred. No.: 88.6 Length: 1197
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 4 Gaps: 0

US-09-712-338-2 (1-555) x US-09-482-273-87 (1-1197)

QY 423 LeuProValArgValSerLeuIle 430
DB 354 CTCCTGTCCGTGTGTCCTTGATC 377

Search completed: November 21, 2003, 22:43:29
Job time : 139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 18:14:28 ; Search time 378 seconds
(without alignments)
3963.461 Million cell updates/sec

Title: US-09-712-338-2

Perfect score: 555

Sequence: 1 MRGYEFLSVLPLVAASWALP.....HTQSSVPLPTATSSSSVGMA 555

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5103490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool.p/US09712338/runat_17112003_170557_11690/app.query.fasta_1.711
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=oli.rng -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcp -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CGN 1.1.490 @runat_17112003_170557_11690 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N Geneseq 19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	554	99.8	1662	19	AAV28620	A. oryzae ATCC2038
2	24	4.3	1656	24	ABZ78288	A. niger serine ca
3	24	4.3	3150	24	ABZ78231	A. niger serine ca
4	14	2.5	626	21	AAF12522	Aspergillus oryzae
5	10	1.8	1581	24	ABZ78241	A. niger serine ca
6	10	1.8	2940	24	ABZ78184	A. niger serine ca
7	9	1.6	299	25	ABX87817	Corn ear-derived p
8	9	1.6	678	21	AAF12116	Aspergillus oryzae
9	9	1.6	1579	21	AAC49933	Arabidopsis thalia
10	9	1.6	1642	21	AAC49932	Arabidopsis thalia
11	9	1.6	1665	24	ABZ78283	A. niger serine ca
12	9	1.6	1670	21	AAC37578	A. niger serine ca
13	9	1.6	1872	24	ABZ78289	A. niger serine ca
14	9	1.6	2027	24	ABQ76315	S. cerevisiae BAX-
15	9	1.6	3080	24	ABZ78226	A. niger serine ca
16	9	1.6	3213	25	ABZ24255	A. niger serine ca
17	9	1.6	3221	24	ABZ78232	Human SLC22A gene
18	9	1.6	3229	23	AAZ75476	A. niger serine ca
19	9	1.6	3229	23	AAZ75478	DNA encoding novel
20	9	1.6	26309	23	AAZ59535	Propionibacterium
21	8	1.4	173	25	ABX29602	Human GDP-mannose
22	8	1.4	223	24	ABN78934	Human ORF3881 cDNA
23	8	1.4	228	22	ABA76502	Human foetal liver
24	8	1.4	228	22	AAK25148	Human brain expres
25	8	1.4	228	22	AAK51154	Human bone marrow
26	8	1.4	228	22	AAI57190	Probe #25876 used
27	8	1.4	228	23	ABS50714	Human liver single
28	8	1.4	228	23	ABS24674	Human genome-deriv
29	8	1.4	269	21	AAC00571	Human secreted pro
30	8	1.4	280	23	ABX50234	Human prostate exp
31	8	1.4	296	25	ABX88328	Corn ear-derived p
32	8	1.4	322	24	ABN75297	Human ORF244 cDNA
33	8	1.4	350	24	ABN18976	Human OREF polynuc
34	8	1.4	370	24	ABV88745	Human colon cancer
35	8	1.4	370	24	ABV89186	Human colon cancer
36	8	1.4	372	22	AAZ35285	Human cardiovascular
37	8	1.4	372	22	AAI62779	Human cDNA SEQ ID
38	8	1.4	372	22	ABN26249	Human OREF polynuc
39	8	1.4	400	22	AAI86967	Human polynucleoti
40	8	1.4	462	24	ABQ67537	Listeria innocua D
41	8	1.4	481	25	ABX22449	Human GDP-mannose
42	8	1.4	495	21	ABN74543	Human OREF ORF98 p
43	8	1.4	495	21	ABN23932	Human OREF polynuc
44	8	1.4	497	25	ABX12491	cDNA encoding corn
45	8	1.4	512	24	ABL79159	Human ovarian canc

ALIGNMENTS

RESULT 1

AAV28620
ID AAV28620 standard; DNA; 1662 BP.

AC AAV28620;

DT 27-AUG-1998 (first entry)

DE A. oryzae ATCC20386 carboxypeptidase I DNA.

KW Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous;
KW food industry; ss.

OS Aspergillus oryzae.

XX Key Location/Qualifiers

PH 1..1662

FT CDS

FT /*tag= a
FT /product= carboxypeptidase I
XX /note= "partial coding sequence"
XX
PN W09814599-A1.
XX
PD 09-APR-1998.
XX
PF 03-OCT-1997; 9TWO-US17977.
XX
XX 27-NOV-1996; 96US-0757534.
PR 04-OCT-1996; 96US-0726880.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO-NORDISK AS.
XX
PI Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;
PI Klotz A, Mathisen TE, Rey M;
XX
DR WPI: 1998-240098/21.
DR P-PSDB; AAW56099.
XX
XX Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous
PT substrates, useful for improving flavour of foods
XX
XX Claim 2; Fig 3; 82pp; English.
XX
CC This DNA sequence encodes carboxypeptidase I from Aspergillus oryzae.
CC This polypeptide has an optimal activity in the range of pH 3.0-7.5 at
CC 25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a
CC residual activity of at least 65.5% after 30 minutes at pH 4.0 and
CC 60 deg. C. It also has the capacity to hydrolyse X from N-CBZ-Ala-X where
CC N-CBZ is N-carboxybenzoyl and X is any amino acid. The carboxypeptidases
CC can be used for obtaining hydrolysates (which can be enriched in free
CC glutamic acid or peptide bound glutamic acid residues) from proteinaceous
CC substrates. The carboxypeptidases can be used in flavour-improving
CC compositions in the food industry. The products can also be used for the
CC production of polypeptides free of carboxypeptidase activity.
XX
SQ Sequence 1662 BP; 396 A; 468 C; 398 G; 400 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 1662
Score: 554.00 Matches: 554
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.82% Indels: 0
DB: 19 Gaps: 0

US-09-712-338-2 (1-555) x AAV28620 (1-1662)

QY	1	MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTrpAlaLeuPro	20
DB	1	ATGCGTGGCTACGAAATTCCTCAGTGTCTACCCCTGGTTGGAGCCAGTTGGGCCCTTCCA	60
QY	21	GlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGlyValLys	40
DB	61	GGAGTACACGGCGTCCGTCGGTAGAAGACAGCTACCCCAAGAACCCACCGGGGTCAAG	120
QY	41	ThrLeuThrThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal	60
DB	121	ACTCTTACACCGCAACATGTCCACATCCGTTACAGGAACCCGGGGAGAGGGGCTC	180
QY	61	CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer	80
DB	181	TGCAGACTACCCGGGTGTCAATCTCTCTGATATGTCGACACTCTCCCGAGTCC	240
QY	81	HisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrIleu	100
DB	241	CATACCTCTTCTGCTTCTCGAAGCCAGACATAACCCAGAACTGCACCTTATCACATG	300
QY	101	TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluGlyPro	120

DB	301	TGGTTGAATGTTGGCCCTGGAGCGATTCTTTGATCGGTCTCTTCCGAAGAGTTGGCCCT	360
QY	121	CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSer	140
DB	361	TGCCATGTCAATTCGACTTTTGTGATGACTACATCAACCTCTCTGTAACGAGTCTCC	420
QY	141	AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp	160
DB	421	AATTTACTATTCTCTGCCAGCCATTGGGAGTCGGCTTTTATATGATGATACGGTTGAT	480
QY	161	GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly	180
DB	481	GGTCCATTAACTTGTAACTGGGGTCTCGAAATTCGAGCTTTCGAGGAGTTCAGGGC	540
QY	181	ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla	200
DB	541	CGGTACCCCAACCATTCGATCGACATACCAATCTTCGCGCAGAGCCGCT	600
QY	201	TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer	220
DB	601	TGGGAGATCCTCGAAGGATTCTTGTAGTGACTTACCTAGCTGGACTCTAGGGTGCAGTCT	660
QY	221	LysAspPheSerLeuTyrThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn	240
DB	661	AAGGACTTCAGTCTATGACGAGAGCTATGGGGCCTATGGTCTCTGCATTTCTCAAT	720
QY	241	HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn	260
DB	721	CATTTTACGACGAGAAATGCAACGCTAGTGTAAATGGTGTTCAGCTTAAT	780
QY	261	PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr	280
DB	781	TTCAACTCTCTGGGAATTAATTAACGCATCTACGACAGCGCATCCAGGCCCTTACTAC	840
QY	281	ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn	300
DB	841	CCTGATTCGCTGTGAACATACCTACGCTATCAGGCTGTCAACGAGACCGTCTCAAC	900
QY	301	TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys	320
DB	901	TACATGAAGTTTGCCAAACCAATGCTGCGAGATTGTGATTTCCACTCGAA	960
QY	321	GlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys	340
DB	961	CAGACAAACCGCACCGCATTTAGCTGACTACGCCCTCTGCCCGAAGCCCAACATGTC	1020
QY	341	ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg	360
DB	1021	AGGACAAATGTTGAGGGGCATACTACGCCCTTTGCTGCTGTGTATGATATTCGG	1080
QY	361	HisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSer	380
DB	1081	CATCCATATGATGACCGACTCCGCCAAGTATTACAAACAAATTTCTGGCAAGGACTCT	1140
QY	381	ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr	400
DB	1141	GTCAIGAACCTATTCGGCTCAACATCACTACCCAGTCCATAATATGACGTCTACTAC	1200
QY	401	AlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle	420
DB	1201	GC'TTTCAGCAACACAGCGGACTTTGTCTGCCCAACTTCATCGAAGACCTCGAGGATC	1260
QY	421	LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp	440
DB	1261	CTTGCTCTCCCGGTGCTCTCCCTCATCTATGGCGACCGCGATACATCTCAACTGG	1320
QY	441	PheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPheArgSer	460
DB	1321	TTCCGCGGTGAGCGCGTTTCCCTCGCTCGCAACTACTCCCAAGCGCGCCAGTCCGAAGC	1380
QY	461	AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly	480
DB	1381	GCAGGTACACGCCCTGAAAGTCAACGGCTCGAGTATGGGGAACCTCGCGAGTATGTT	1440

Qy 481 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProile 500
 Db 1441 AATTCTCTCTCACTCGCTCTATGAGCGCCATGAAGTCCCATCTACTACGACCCATC 1500
 Qy 501 AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys 520
 Db 1501 GCCTCCCTGCAATGTTTAAACCGACTATCTTCGGTTGGATATCGCAGAGGGCCAGAAG 1560
 Qy 521 LysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 540
 Db 1561 AAGATCTGGCCAGCTACAGACCAATGGAACGGCTACAGCTACGCATACACACATCGTCC 1620
 Qy 541 ValProLeuProThrAlaThrSerMetSerSerValGlyMet 554
 Db 1621 GTGCCGCTGCTACGGCTACCGATCCAGCATGTCCAGTGTGGTATG 1662

RESULT 2
 ABZ78288
 ID ABZ78288 standard; cDNA; 1656 BP.
 XX AC ABZ78288;
 XX DT 24-APR-2003 (first entry)
 XX DE A. niger serine carboxypeptidase cDNA #3.
 XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ss.
 XX OS Aspergillus niger.
 XX FH Key Location/Qualifiers
 FT CDS 1..1656
 FT /*tag= a
 FT /EC_number= "3.4.16.6"
 XX WO200268623-A2.
 XX PD 06-SEP-2002.
 XX XX 22-FEB-2002; 2002WO-EP01984.
 XX 23-FEB-2001; 2001EP-0200657.
 PR 23-FEB-2001; 2001EP-0200658.
 PR 23-FEB-2001; 2001EP-0200660.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 28-MAR-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0200075.
 PR 28-MAR-2001; 2001EP-0200078.
 PR 28-MAR-2001; 2001EP-0200080.
 PR 28-MAR-2001; 2001EP-0200087.
 PR 28-MAR-2001; 2001EP-0200088.
 PR 21-MAY-2001; 2001EP-0200156.
 PR 21-MAY-2001; 2001EP-0200159.
 PR 21-MAY-2001; 2001EP-0200160.
 PR 21-MAY-2001; 2001EP-0200162.
 PR 21-MAY-2001; 2001EP-0200165.
 PR 21-MAY-2001; 2001EP-0200166.
 PR 21-MAY-2001; 2001EP-0200168.
 PR 21-JUN-2001; 2001EP-0200240.
 PR 21-JUN-2001; 2001EP-0200242.
 PR 21-JUN-2001; 2001EP-0200244.
 PR 21-JUN-2001; 2001EP-0200246.
 PR 12-JUL-2001; 2001EP-0200280.
 PR 12-JUL-2001; 2001EP-0200285.
 PR 30-JUL-2001; 2001EP-0200323.
 PR 30-JUL-2001; 2001EP-0200327.
 PR 02-AUG-2001; 2001EP-0200341.
 PR 02-AUG-2001; 2001EP-0200342.

PR 02-AUG-2001; 2001EP-0200343.
 PR 02-AUG-2001; 2001EP-0200344.
 PR 09-AUG-2001; 2001EP-0200357.
 PR 16-AUG-2001; 2001EP-0200374.
 PR 16-AUG-2001; 2001EP-0200377.
 PR 20-SEP-2001; 2001EP-0200478.
 PR 20-SEP-2001; 2001EP-0200483.
 PR 22-OCT-2001; 2001EP-0200552.
 PR 22-OCT-2001; 2001EP-0200553.
 PR 22-OCT-2001; 2001EP-0200554.
 PR 22-OCT-2001; 2001EP-0200556.
 PR 22-OCT-2001; 2001EP-0200557.
 PR 22-OCT-2001; 2001EP-0200558.
 PR 15-NOV-2001; 2001EP-0200464.
 PR 21-DEC-2001; 2001EP-02005117.
 XX (STAM) DSM NV.
 PA Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 XX Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Maier D, Spredafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Stiebler J, Albang R;
 XX WPI; 2002-723203/78.
 DR P-PSDB; ABR38864.
 DR XX Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications
 XX Claim 1; Page 252-253; 394pp; English.
 XX CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 CC A. niger proteases of the invention.
 XX SQ Sequence 1656 BP; 411 A; 426 C; 409 G; 410 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.08e-14 Length: 1656
 Score: 24.00 Matches: 24
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.32% Indels: 0
 DB: 24 Gaps: 0
 US-09-712-338-2 (1-555) x ABZ78288 (1-1656)
 Qy 99 ThrLeuTyrLeuAsnGlyGlyProGlySerLeuLeuGlyLeuPheGluGluLeu 118
 Db 295 ACTCTGTGCTGAATGGTGGCCCTCGAAGCGGATTCCTTGATGGCTTTTGAAGATTG 354
 Qy 119 GlyProCysHis 122
 Db 355 GGTCCGTGTCAC 366
 RESULT 3
 ABZ78231
 ID ABZ78231 standard; DNA; 3150 BP.
 XX AC ABZ78231;
 XX XX

DT 24-APR-2003 (first entry)
 XX A. niger serine carboxypeptidase gene #3.
 DE
 XX
 KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW Gene; ds.
 XX Aspergillus niger.
 XX WO200268623-A2.
 PN
 XX
 PD 06-SEP-2002.
 XX
 XX 22-FEB-2002; 2002WO-EP01984.
 PF
 XX 23-FEB-2001; 2001EP-0200657.
 PR 23-FEB-2001; 2001EP-0200658.
 PR 23-FEB-2001; 2001EP-0200660.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 26-FEB-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0000075.
 PR 28-MAR-2001; 2001EP-0000078.
 PR 28-MAR-2001; 2001EP-0000080.
 PR 28-MAR-2001; 2001EP-0000087.
 PR 28-MAR-2001; 2001EP-0000088.
 PR 21-MAY-2001; 2001EP-0000156.
 PR 21-MAY-2001; 2001EP-0000159.
 PR 21-MAY-2001; 2001EP-0000160.
 PR 21-MAY-2001; 2001EP-0000162.
 PR 21-MAY-2001; 2001EP-0000165.
 PR 21-MAY-2001; 2001EP-0000166.
 PR 21-MAY-2001; 2001EP-0000168.
 PR 21-JUN-2001; 2001EP-0000240.
 PR 21-JUN-2001; 2001EP-0000242.
 PR 21-JUN-2001; 2001EP-0000244.
 PR 21-JUN-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 12-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 02-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 15-NOV-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0004464.
 PR 21-DEC-2001; 2001EP-0005117.
 XX
 XX (STAM) DSM NV.
 XX
 XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Srebleier J, Albang R;
 XX
 XX WPI; 2002-723203/78.
 DR P-PSDB; ABR38864.
 DR
 XX Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT

PT applications
 XX Claim 1; Page 187-189; 394pp; English.
 PS
 XX
 CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderizing), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the
 CC A. niger proteases of the invention.
 XX
 SQ Sequence 3150 BP; 801 A; 778 C; 738 G; 833 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.99e-14 Length: 3150
 Score: 24.00 Matches: 24
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.32% Indels: 0
 DB: 24 Gaps: 0
 US-09-712-338-2 (1-555) x ABZ78231 (1-3150)
 QY 398 ValTyrTyrAlaPheGlnGlnThrGlyAspPheValTyrProAsnPhelleGluAspLeu 417
 DB 2295 GTATTATTGTCATTCAGACACCGCGACTTTGTAIGGCCGAATTCATTGAGGACCTC 2354
 QY 418 GluGluIleLeu 421
 DB 2355 GAAGAGATCCTC 2366
 RESULT 4
 ID AAF12522 standard; cDNA; 626 BP.
 XX
 AC AAF12522;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:5045.
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 XX WO200056762-A2.
 PN
 XX
 PD 28-SEP-2000.
 XX
 XX 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI; 2000-594572/56.
 XX
 XX Monitoring differential expression of genes in filamentous fungal cells
 PT

PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 XX substrate of expressed sequence tags -
 PS Claim 88; Page 2118; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of the
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF1247 represents ESTs from
 CC *Fusarium venenatum*; AAF1248 to AAF1853 represents ESTs from *Aspergillus*
 CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
 CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
 CC all specifically claimed in the present invention.

SX Sequence 626 BP; 153 A; 175 C; 145 G; 153 T; 0 other;

Alignment Scores:

Pred. No.: 0.000111 Length: 626
 Score: 14.00 Matches: 14
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.52% Indels: 0
 DB: 21 Gaps: 0

US-09-712-338-2 (1-555) x AAF12522 (1-626)

QY 61 CysGluThrProGlyValLysSerTyrSerGlyTyrVal 74
 DB 181 TGGCAACACCCAGCGTGCAAACTGCTACTCGGTACGTG 222

RESULT 5

ID ABZ78241 standard; cDNA; 1581 BP.
 XX AC ABZ78241;

XX 24-APR-2003 (first entry)

DE A. niger serine carboxypeptidase cDNA #1.

XX Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ss.
 XX *Aspergillus niger*.

XX Key Location/Qualifiers
 FH 1..1581
 FT CDS /*tag= a
 FT /BC_number= "3.4.16.6"

XX WO200268623-A2.

XX 06-SEP-2002.

XX 22-FEB-2002; 2002WO-EP01984.

PR 23-FEB-2001; 2001EP-0200657.
 PR 23-FEB-2001; 2001EP-0200658.
 PR 23-FEB-2001; 2001EP-0200660.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 26-FEB-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0000075.
 PR 28-MAR-2001; 2001EP-0000078.
 PR 28-MAR-2001; 2001EP-0000080.
 PR 28-MAR-2001; 2001EP-0000087.
 PR 28-MAR-2001; 2001EP-0000088.
 PR 21-MAY-2001; 2001EP-0000156.
 PR 21-MAY-2001; 2001EP-0000159.
 PR 21-MAY-2001; 2001EP-0000160.
 PR 21-MAY-2001; 2001EP-0000162.
 PR 21-MAY-2001; 2001EP-0000165.
 PR 21-MAY-2001; 2001EP-0000166.
 PR 21-MAY-2001; 2001EP-0000168.
 PR 21-JUN-2001; 2001EP-0000240.
 PR 21-JUN-2001; 2001EP-0000242.
 PR 21-JUN-2001; 2001EP-0000244.
 PR 12-JUL-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 30-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000323.
 PR 02-AUG-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 15-NOV-2001; 2001EP-0000558.
 PR 21-DEC-2001; 2001EP-0000464.
 PR 21-DEC-2001; 2001EP-0000517.

(STAM) DSM NV.

Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 Stiebler J, Albarg R;

WPI; 2002-723203/78.

P-PSDB; ABR38817.

Novel isolated protease polypeptide useful in laboratory, clinical,
 pharmaceutical, chemical, diagnostic, personal care and industrial
 applications -

Claim 1; Page 204-205; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A
 polypeptide or polynucleotide of the invention is useful for diagnosing a
 fungal infection such as aspergillosis, or as a query sequence to perform
 a search against public databases. A polypeptide of the invention is
 useful in a selected number of industrial or pharmaceutical processes, in
 laboratory or clinical processes, in food industry (baking, brewing,
 cheese manufacturing, meat tenderising), in tanning industry and in the
 manufacture of biological detergents. A polypeptide may also be useful
 for improving protein solubility, extraction yields, viscosity or taste,
 texture, nutritional value, minimising of antigenicity or
 anti-nutritional factors, colour or functionality as well as processing
 aspects like filterability of the proteinaceous raw material. The

CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 CC A. niger proteases of the invention.

XX Sequence 1581 BP; 343 A; 473 C; 423 G; 342 T; 0 other;

Alignment Scores:
 Pred. No.: 3.91 Length: 1581
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.80% Indels: 0
 DB: 24 Gaps: 0

US-09-712-338-2 (1-555) x ABZ78241 (1-1581)

QY 227 ThrGluSerTyrGlyHisTyrGlyPro 236
 DB 541 ACGGAGAGTTATGGCGGCCTACGGGCC 570

RESULT 6

ID ABZ78184 standard; DNA; 2940 BP.

XX AC ABZ78184;

XX DT 24-APR-2003 (first entry)

XX A. niger serine carboxypeptidase gene #1.

XX Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ds.

XX Aspergillus niger.

XX WO200268623-A2.

XX 06-SEP-2002.

XX 22-FEB-2002; 2002WO-EP01984.

XX 23-FEB-2001; 2001EP-0200657.

XX 23-FEB-2001; 2001EP-0200658.

XX 26-FEB-2001; 2001EP-0200660.

XX 26-FEB-2001; 2001EP-0200706.

XX 26-FEB-2001; 2001EP-0200707.

XX 26-FEB-2001; 2001EP-0200708.

XX 26-FEB-2001; 2001EP-0200719.

XX 28-MAR-2001; 2001EP-0000075.

XX 28-MAR-2001; 2001EP-0000078.

XX 28-MAR-2001; 2001EP-0000080.

XX 28-MAR-2001; 2001EP-0000087.

XX 28-MAR-2001; 2001EP-0000088.

XX 21-MAY-2001; 2001EP-0000156.

XX 21-MAY-2001; 2001EP-0000159.

XX 21-MAY-2001; 2001EP-0000160.

XX 21-MAY-2001; 2001EP-0000162.

XX 21-MAY-2001; 2001EP-0000165.

XX 21-MAY-2001; 2001EP-0000166.

XX 21-MAY-2001; 2001EP-0000168.

XX 21-JUN-2001; 2001EP-0000240.

XX 21-JUN-2001; 2001EP-0000242.

XX 21-JUN-2001; 2001EP-0000244.

XX 21-JUN-2001; 2001EP-0000246.

XX 12-JUL-2001; 2001EP-0000280.

XX 12-JUL-2001; 2001EP-0000285.

XX 30-JUL-2001; 2001EP-0000323.

XX 30-JUL-2001; 2001EP-0000327.

XX 02-AUG-2001; 2001EP-0000341.

XX 02-AUG-2001; 2001EP-0000342.

XX 02-AUG-2001; 2001EP-0000343.

XX 02-AUG-2001; 2001EP-0000344.

XX 09-AUG-2001; 2001EP-0000357.

PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000479.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 22-OCT-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0000464.
 PR 21-DEC-2001; 2001EP-0005117.
 XX

(STAM) DSM NV.

XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Stiebler J, Albang R;

XX WPI; 2002-723203/78.

DR P-PSDB; ABR38817.

XX Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications -

XX Claim 1; Page 109-111; 394pp; English.

XX The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the
 CC A. niger proteases of the invention.

XX Sequence 2940 BP; 691 A; 818 C; 732 G; 699 T; 0 other;

Alignment Scores:

Pred. No.:	7.07	Length:	2940
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.80%	Indels:	0
DB:	24	Gaps:	0

US-09-712-338-2 (1-555) x ABZ78184 (1-2940)

QY 227 ThrGluSerTyrGlyHisTyrGlyPro 236

DB 1163 ACGGAGAGTTATGGCGGCCTACGGGCC 1192

RESULT 7

ABZ87817

ID ABZ87817 standard; cDNA; 299 BP.

XX ABZ87817;

XX 24-APR-2003 (first entry)

DE Corn ear-derived polynucleotide (cpd) #6277.

XX Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022;

KW SATMON023; structural gene; functional gene; regulatory gene;

KW corn ear-specific profile; gene transcription; gene expression;

KW hybrid plant; desirable trait expression; plant breeding program;
 KW inheritance; desired characteristic; growth; development;
 KW disease resistance; environmental adaptability; quality; yield;
 KW multigene trait; plant; gene; ss.
 XX
 OS Zea mays.
 XX
 XX US6476212-B1.
 PN
 XX
 PD 05-NOV-2002.
 XX
 XX 14-MAY-1999; 99US-0313294.
 PF
 XX 26-MAY-1998; 99US-086722P.
 PR
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Lalgudi RV, Ito LY, Sherman BK;
 PI
 XX WPI; 2003-208840/20.
 DR
 XX Novel purified corn-ear derived polynucleotide useful as hybridization
 PT probe for detecting polynucleotide in sample, and for identifying,
 PT evaluating, and altering desired characteristics associated with
 PT growth, development
 PT
 XX Examples; SEQ ID No 6277; 390pp; English.
 PS
 XX The present invention relates to the isolation of corn ear-derived
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries
 CC SATMON022 and SATMON023. Some of the cpds uniquely identify structural,
 CC functional, and regulatory genes of corn ear. The polynucleotides
 CC sequences are useful for detecting cpds in a sample, for producing
 CC a corn ear-specific profile of gene transcription, for detecting
 CC altered gene expression in inbred or hybrid plants, and for screening
 CC several molecules for specific binding to the polynucleotide. The cpds
 CC are useful to identify, isolate, or extend identical or related
 CC corn-ear nucleic acid sequences from DNA libraries, and in nucleic
 CC acid amplification or hybridisation techniques to follow the
 CC expression of desirable traits through plant breeding programs.
 CC Preferably, the cpds are used to identify, evaluate, alter, or
 CC follow the inheritance of desired characteristics associated with
 CC growth and development, disease resistance, environmental adaptability,
 CC quality and yield of corn. The cpds are also useful as molecular
 CC markers for studying inheritance and multigene traits in a plant
 CC breeding program. The cpds are useful for producing purified corn-ear
 CC polypeptides by recombinant techniques. They are also useful in
 CC diagnostic assays to detect or confirm conditions or diseases
 CC associated with abnormal levels of cdp expression. ABX81541-ABX89140
 CC represent corn ear-derived polynucleotides (cpds) of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site at seqdata.uspto.gov/psipspIDEntry.html.
 XX
 SQ Sequence 299 BP; 65 A; 102 C; 73 G; 59 T; 0 other;
 Alignment Scores:
 Pred. No.: 8.8 Length: 299
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 1.62%
 Indels: 0
 DB: 25 Gaps: 0

US-09-712-338-2 (1-555) x ABX87817 (1-299)

OY 540 SerValProLeuProThrAlaThrSer 548
 Db |||||
 48 TCTGTGCCACTGCCACCTCA 74

RESULT 8
 AAF12116
 ID AAF12116 standard; cDNA; 678 BP.

KW hybrid plant; desirable trait expression; plant breeding program;
 KW inheritance; desired characteristic; growth; development;
 KW disease resistance; environmental adaptability; quality; yield;
 KW multigene trait; plant; gene; ss.
 XX
 OS Zea mays.
 XX
 XX US6476212-B1.
 PN
 XX
 PD 05-NOV-2002.
 XX
 XX 14-MAY-1999; 99US-0313294.
 PF
 XX 26-MAY-1998; 99US-086722P.
 PR
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Lalgudi RV, Ito LY, Sherman BK;
 PI
 XX WPI; 2003-208840/20.
 DR
 XX Novel purified corn-ear derived polynucleotide useful as hybridization
 PT probe for detecting polynucleotide in sample, and for identifying,
 PT evaluating, and altering desired characteristics associated with
 PT growth, development
 PT
 XX Examples; SEQ ID No 6277; 390pp; English.
 PS
 XX The present invention relates to the isolation of corn ear-derived
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries
 CC SATMON022 and SATMON023. Some of the cpds uniquely identify structural,
 CC functional, and regulatory genes of corn ear. The polynucleotides
 CC sequences are useful for detecting cpds in a sample, for producing
 CC a corn ear-specific profile of gene transcription, for detecting
 CC altered gene expression in inbred or hybrid plants, and for screening
 CC several molecules for specific binding to the polynucleotide. The cpds
 CC are useful to identify, isolate, or extend identical or related
 CC corn-ear nucleic acid sequences from DNA libraries, and in nucleic
 CC acid amplification or hybridisation techniques to follow the
 CC expression of desirable traits through plant breeding programs.
 CC Preferably, the cpds are used to identify, evaluate, alter, or
 CC follow the inheritance of desired characteristics associated with
 CC growth and development, disease resistance, environmental adaptability,
 CC quality and yield of corn. The cpds are also useful as molecular
 CC markers for studying inheritance and multigene traits in a plant
 CC breeding program. The cpds are useful for producing purified corn-ear
 CC polypeptides by recombinant techniques. They are also useful in
 CC diagnostic assays to detect or confirm conditions or diseases
 CC associated with abnormal levels of cdp expression. ABX81541-ABX89140
 CC represent corn ear-derived polynucleotides (cpds) of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site at seqdata.uspto.gov/psipspIDEntry.html.
 XX
 SQ Sequence 299 BP; 65 A; 102 C; 73 G; 59 T; 0 other;
 Alignment Scores:
 Pred. No.: 8.8 Length: 299
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 1.62%
 Indels: 0
 DB: 25 Gaps: 0

US-09-712-338-2 (1-555) x ABX87817 (1-299)

OY 540 SerValProLeuProThrAlaThrSer 548
 Db |||||
 48 TCTGTGCCACTGCCACCTCA 74

RESULT 8
 AAF12116
 ID AAF12116 standard; cDNA; 678 BP.

XX
 AC AAF12116;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:4639.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 XX 22-MAR-2000; 2000WO-US07781.
 PF
 XX 22-MAR-1999; 99US-0273623.
 PR
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 XX
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 XX WPI; 2000-594572/56.
 DR
 XX Monitoring differential expression of genes in filamentous fungal cells
 CC uses fluorescence-labeled nucleic acids isolated from the cells and a
 CC substrate of expressed sequence tags
 CC
 XX Claim 88; Page 1971; 3161pp; English.
 PS
 XX The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 678 BP; 153 A; 188 C; 175 G; 161 T; 1 other;
 Alignment Scores:
 Pred. No.: 19.2 Length: 678
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 1.62%
 Indels: 0
 DB: 21 Gaps: 0

US-09-712-338-2 (1-555) x AAF12116 (1-678)

Mon Nov 24 13:41:06 2003

QY 479 TyrGlyAenPheSerPheThrArgVal 487
Db 395 TACGGAACCTCTCGTTCACCTCGGTT 421

RESULT 9
ID AAC49933 standard; DNA; 1579 BP.
XX AAC49933;
AC AAC49933;
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 62983.
XX Arabidopsis thaliana.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136332.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

```

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Alignment Scores:
Pred. No.: 43
Score: 9.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.62%
DB: 21
Length: 1579
Matches: 9
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```

US-09-712-338-2 (1-555) x AAC49933 (1-1579)

QY 99 ThrLeuTrpIleuAnsclyGlyProGly 107
|||||

```

Db 232 ACCCTTTGGCTCAATGAGGTCCAGGT 258
RESULT 10
AAC49932
ID AAC49932 standard; DNA; 1642 BP.
XX AAC49932;
AC AAC49932;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 62979.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EPI033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.

```

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:

Pred. No.:	44	6	Length:	1642
Score:	9.00		Matches:	9
Percent Similarity:	100.00%		Conservative:	0
Best Local Similarity:	100.00%		Mismatches:	0
Query Match:	1.62%		Indels:	0
DB:	21		Gaps:	0

US-09-712-338-2 (1-555) x AAC49932 (1-1642)

Qy 99 ThrLeuTrpLeuMetGlyProGly 107

Db 279 ACCCTTTGGCTCAATGGAGGTCCAGT 305

RESULT 11
 ABZ78283
 ID ABZ78283 standard; cDNA; 1665 BP.
 AC ABZ78283;
 XX
 DT 24-APR-2003 (first entry)
 XX
 DE A. niger serine carboxypeptidase cDNA #2.
 XX
 KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ss.
 XX
 OS Aspergillus niger.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1665
 FT /*tag= a
 FT /EC_number= "3.4.16.6"
 XX
 PN WO200268623-A2.
 XX
 PD 06-SEP-2002.
 XX
 FF 22-FEB-2002; 2002WO-EP01984.
 XX
 PR 23-FEB-2001; 2001EP-0200657.
 PR 23-FEB-2001; 2001EP-0200658.
 PR 23-FEB-2001; 2001EP-0200660.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 26-FEB-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0000075.
 PR 28-MAR-2001; 2001EP-0000078.
 PR 28-MAR-2001; 2001EP-0000080.
 PR 28-MAR-2001; 2001EP-0000087.
 PR 28-MAR-2001; 2001EP-0000088.
 PR 21-MAY-2001; 2001EP-0000156.
 PR 21-MAY-2001; 2001EP-0000159.
 PR 21-MAY-2001; 2001EP-0000160.
 PR 21-MAY-2001; 2001EP-0000162.
 PR 21-MAY-2001; 2001EP-0000165.
 PR 21-MAY-2001; 2001EP-0000166.
 PR 21-MAY-2001; 2001EP-0000168.
 PR 21-JUN-2001; 2001EP-0000240.
 PR 21-JUN-2001; 2001EP-0000242.
 PR 21-JUN-2001; 2001EP-0000244.
 PR 21-JUN-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 12-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000323.
 PR 30-JUL-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 15-NOV-2001; 2001EP-0000558.
 PR 21-DEC-2001; 2001EP-0004464.
 PR 21-DEC-2001; 2001EP-0005117.
 XX
 PA (STAM) DSM NV.

XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Mater D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Stiebler J, Albang R;
 XX
 DR WPI: 2002-723203/78.
 DR P-PSDB; ABR38859.
 XX
 PT Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications
 XX
 PS Claim 1; Page 246-247; 394pp; English.
 XX
 CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 CC A. niger proteases of the invention.
 XX
 SQ Sequence 1665 BP; 374 A; 455 C; 424 G; 412 T; 0 other;
 Alignment Scores:
 Pred. No.: 45.2 Length: 1665
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.62% Indels: 0
 DB: 24 Gaps: 0
 US-09-712-338-2 (1-555) x ABZ78283 (1-1665)
 QY 228 GluSerTyrGlyHisTyrGlyPro 236
 DB 559 GAAAGTTACGGAGCCATTATGTCCTCA 585
 RESULT 12
 AAC37578
 ID AAC37578 standard; DNA; 1670 BP.
 XX
 AC AAC37578;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 17899.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.

us-09-712-338-2.oli.rng

Mon Nov 24 13:41:06 2003

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 15-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
 Pred. No.: 45.4
 Score: 9.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 1.62%
 DB: 21
 Length: 1670
 Matches: 9
 Conservative: 0
 Mismatches: 0
 Indels: 0
 Gaps: 0

US-09-712-338-2 (1-555) x AAC37578 (1-1670)

QY 99 ThrLeuTrpLeuAsnGlyProGly 107
 |||||
 Db 278 ACCCTTGGCTTAATGAGGCTCTGGT 304

RESULT 13

ABZ78289
 ID ABZ78289 standard; cDNA; 1872 BP.
 AC ABZ78289;
 XX
 DT 24-APR-2003 (first entry)
 DE A. niger serine carboxypeptidase cDNA #4.
 KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ss.
 XX
 OS Aspergillus niger.

XX
 FH Key Location/Qualifiers
 FT CDS 1..1872
 FT /*tag= a
 FT /EC_number= "3.4.16.6"

XX WO200268623-A2.

XX PD 06-SEP-2002.

XX PF 22-FEB-2002; 2002WO-EP01984.

XX XX 23-FEB-2001; 2001EP-0200657.

PR 23-FEB-2001; 2001EP-0200658.

PR 23-FEB-2001; 2001EP-0200660.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 26-FEB-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0000075.
 PR 28-MAR-2001; 2001EP-0000078.
 PR 28-MAR-2001; 2001EP-0000080.
 PR 28-MAR-2001; 2001EP-0000087.
 PR 28-MAR-2001; 2001EP-0000088.
 PR 21-MAY-2001; 2001EP-0000156.
 PR 21-MAY-2001; 2001EP-0000159.
 PR 21-MAY-2001; 2001EP-0000160.
 PR 21-MAY-2001; 2001EP-0000182.
 PR 21-MAY-2001; 2001EP-0000185.
 PR 21-MAY-2001; 2001EP-0000186.
 PR 21-MAY-2001; 2001EP-0000168.
 PR 21-JUN-2001; 2001EP-0000240.
 PR 21-JUN-2001; 2001EP-0000242.
 PR 21-JUN-2001; 2001EP-0000244.
 PR 21-JUN-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 12-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000323.
 PR 30-JUL-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 22-OCT-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0000464.
 PR 21-DEC-2001; 2001EP-00005117.

(STAM) DSM NV.

Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel B;
 Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 Stiebler J, Albang R;

WPI; 2002-723203/79.

P-PSDH; ABR38865.

Novel isolated protease polypeptide useful in laboratory, clinical,
 pharmaceutical, chemical, diagnostic, personal care and industrial
 applications

Claim 1; Page 253-254; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A
 polypeptide or polynucleotide of the invention is useful for diagnosing a
 fungal infection such as aspergillosis, or as a query sequence to perform
 a search against public databases. A polypeptide of the invention is
 useful in a selected number of industrial or pharmaceutical processes, in
 laboratory or clinical processes, in food industry (baking, brewing,
 cheese manufacturing, meat tenderising), in tanning industry and in the
 manufacture of biological detergents. A polypeptide may also be useful
 for improving protein solubility, extraction yields, viscosity or taste,
 texture, nutritional value, minimising of antigenicity or
 anti-nutritional factors, colour or functionality as well as processing
 aspects like filterability of the proteinaceous raw material. The
 sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 A. niger proteases of the invention.

XX SQ Sequence 1872 BP; 402 A; 434 C; 570 G; 466 T; 0 other;
Alignment Scores:
Pred. No.: 50.6 Length: 1872
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 24 Gaps: 0

US-09-712-338-2 (1-555) x ABZ78289 (1-1872)

QY 479 TyrGlyAsnPheserPheThrArgVal 487
DB 1492 TATGGAACTTCTCGTTCACGGGTGTC 1518

RESULT 14
ABQ76315
ID ABQ76315 standard; cDNA; 2027 BP.

XX AC ABQ76315;
XX DT 21-NOV-2002 (first entry)
XX DE S. cerevisiae BAX-associated cDNA fragment SEQ ID 55.

XX Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
KW neurodegeneration; cell death; ss.

XX OS Saccharomyces cerevisiae.

XX FN WC00264766-A2.

XX PD 22-AUG-2002.

XX PF 21-DEC-2001; 2001WO-EP15398.

XX PR 22-DEC-2000; 2000EP-0870318.

XX PR 04-JAN-2001; 2001EP-0870002.

XX PR 09-JAN-2001; 2001EP-0870003.

XX PA (JANC) JANSSEN PHARM NV.

XX PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;

XX DR WPI; 2002-667002/71.

XX DR P-PSDB; ABG93049.

XX PT New isolated nucleic acid representing a synthetic Bax gene, useful as
PT medicament for treating, preventing and/or alleviating yeast or fungal
PT infections or proliferative disorders, or for preventing apoptosis in
PT certain diseases -

XX PS Claim 36; Figure 1; 344pp; English.

XX CC This invention describes a novel nucleic acid representing a synthetic
CC Bax gene. The Bax gene of the invention is useful for identifying
CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying
CC Candida spp. sequences that are differentially expressed in a pathway
CC eventually leading to programmed cell death or identifying inhibitors or
CC inhibitor sequences of Bax-induced cell death. The products of the
CC invention have cytostatic, fungicide, immunosuppressive, virucide and
CC vasotropic activity and can be used in vaccines or for gene therapy. The
CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
CC antisense molecules and antibodies are useful as medicaments or in
CC preparing a medicament for treating, preventing and/or alleviating
CC diseases associated with yeast or fungi or proliferative disorders, such
CC as cancer, or for preventing apoptosis in certain diseases. The compounds
CC or polypeptides, or the genetically modified organism are useful for
CC preparing a medicament for modifying the endogenous flora of humans and

CC other mammals. The vaccine is useful for immunising against yeast or
CC fungal infections. Apoptosis-related diseases include autoimmune disease,
CC ischaemia, diseases related with viral infections or neurodegenerations.
CC This sequence represents a polynucleotide associated with the Bax gene
CC described in the disclosure of the invention.

XX SQ Sequence 2027 BP; 583 A; 380 C; 427 G; 637 T; 0 other;

Alignment Scores:
Pred. No.: 54.6 Length: 2027
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 24 Gaps: 0

US-09-712-338-2 (1-555) x ABQ76315 (1-2027)

QY 147 GlnProLeuGlyValGlyPheSerTyr 155
DB 1011 CAGCCACTCGGAGTCGGCTTTTCCTAT 1037

RESULT 15
ABZ78226

ID ABZ78226 standard; DNA; 3080 BP.

XX AC ABZ78226;

XX DT 24-APR-2003 (first entry)

XX DE A. niger serine carboxypeptidase gene #2.

XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW gene; ds.

XX OS Aspergillus niger.

XX PN WC000268623-A2.

XX PD 06-SEP-2002.

XX PF 22-FEB-2002; 2002WO-EP01984.

XX PR 23-FEB-2001; 2001EP-0200657.

XX PR 23-FEB-2001; 2001EP-0200658.

XX PR 23-FEB-2001; 2001EP-0200660.

XX PR 26-FEB-2001; 2001EP-0200706.

XX PR 26-FEB-2001; 2001EP-0200707.

XX PR 26-FEB-2001; 2001EP-0200708.

XX PR 26-FEB-2001; 2001EP-0200719.

XX PR 28-MAR-2001; 2001EP-0000075.

XX PR 28-MAR-2001; 2001EP-0000078.

XX PR 28-MAR-2001; 2001EP-0000080.

XX PR 28-MAR-2001; 2001EP-0000087.

XX PR 28-MAR-2001; 2001EP-0000088.

XX PR 21-MAY-2001; 2001EP-0000156.

XX PR 21-MAY-2001; 2001EP-0000159.

XX PR 21-MAY-2001; 2001EP-0000160.

XX PR 21-MAY-2001; 2001EP-0000162.

XX PR 21-MAY-2001; 2001EP-0000165.

XX PR 21-MAY-2001; 2001EP-0000166.

XX PR 21-MAY-2001; 2001EP-0000168.

XX PR 21-JUN-2001; 2001EP-0000240.

XX PR 21-JUN-2001; 2001EP-0000242.

XX PR 21-JUN-2001; 2001EP-0000244.

XX PR 12-JUN-2001; 2001EP-0000246.

XX PR 12-JUL-2001; 2001EP-0000280.

XX PR 12-JUL-2001; 2001EP-0000285.

XX PR 30-JUL-2001; 2001EP-0000323.

XX PR 30-JUL-2001; 2001EP-0000327.

XX PR 02-AUG-2001; 2001EP-0000341.

XX PR 02-AUG-2001; 2001EP-0000342.

PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000344.
PR 09-AUG-2001; 2001EP-0000357.
PR 16-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.
PR 22-OCT-2001; 2001EP-0000556.
PR 22-OCT-2001; 2001EP-0000557.
PR 22-OCT-2001; 2001EP-0000558.
PR 15-NOV-2001; 2001EP-0004464.
PR 21-DEC-2001; 2001EP-0005117.
XX
XX
PA (STAM) DSM NV.
XX
XX
PI Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albang R;
XX
XX WPI: 2002-723203/78.
DR P-PSDB; ABR38859.
DR
DR
XX
XX
PT Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications .
XX
XX
PS Claim 1; Page 178-180; 394pp; English.
XX
XX
CC The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the
CC A. niger proteases of the invention.
XX
SQ Sequence 3080 BP; 755 A; 803 C; 735 G; 787 T; 0 other;

Alignment Scores:
Pred. No.: 81.3 Length: 3080
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.62% Indels: 0
DB: 24 Gaps: 0

US-09-712-338-2 (1-555) x ABZ78226 (1-3080)

QY 228 GluSerTyrGlyGlyHisTyrGlyPro 236
Db 1334 GAAAGTTCGAGGAGCATTTATGTCCTCA 1360

Search completed: November 21, 2003, 20:35:49
Job time : 391 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 14:28:22 ; Search time 122 Seconds
(without alignments)
6034.648 Million cell updates/sec

Title: US-09-712-338-1
Perfect score: 1668
Sequence: 1 atgcgtggtacgaattctt.....ccagtgttgatgcatag 1668

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B COMB.seq: *
5: /cgn2_6/prodata/2/ina/PTUS COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1662	99.6	1662	3	US-08-943-714-1
2	18	1.1	716	4	US-09-620-312D-1066
3	18	1.1	729	4	US-09-328-352-867
4	17	1.0	513	4	US-09-252-991A-5290
5	17	1.0	540	4	US-09-252-991A-14098
6	17	1.0	768	4	US-09-252-991A-5185
7	17	1.0	1116	4	US-09-221-047B-125
8	17	1.0	1319	2	US-08-504-459-7
9	17	1.0	1347	4	US-09-134-001C-373
10	17	1.0	1629	4	US-09-252-991A-13416
11	17	1.0	1731	4	US-09-252-991A-13732
12	17	1.0	2051	4	US-09-399-913-52
13	17	1.0	2838	4	US-09-252-991A-14128
14	17	1.0	3897	4	US-09-252-991A-14008
15	17	1.0	4558	1	US-08-309-512-2
16	17	1.0	4558	5	PCT-US92-08756A-2
17	17	1.0	4781	2	US-09-001-273-1
18	17	1.0	4781	3	US-08-843-459A-1
19	17	1.0	4847	3	US-09-061-400-1
20	17	1.0	5115	3	US-08-348-518C-3
21	17	1.0	5115	3	US-08-476-509B-3
22	17	1.0	5326	3	US-08-751-359-21
23	17	1.0	5526	4	US-08-907-146-21
24	17	1.0	536165	4	US-09-214-808-1
25	17	1.0	1830121	4	US-09-557-884-1
26	17	1.0	1830121	4	US-09-643-990A-1
27	17	1.0	4403765	3	US-09-103-840A-2

28 17 1.0 4411529 3 US-09-103-840A-1 Sequence 1, Appli
29 16 1.0 303 4 US-09-252-991A-7045 Sequence 7045, Ap
30 16 1.0 333 4 US-09-795-926-3 Sequence 3, Appli
31 16 1.0 357 4 US-09-280-116-131 Sequence 131, Appl
32 16 1.0 359 2 US-08-687-080-65 Sequence 65, Appl
33 16 1.0 372 4 US-09-221-017B-258 Sequence 258, App
34 16 1.0 391 4 US-09-389-681-245 Sequence 245, App
35 16 1.0 391 4 US-09-620-405B-245 Sequence 245, App
36 16 1.0 391 4 US-09-339-338-245 Sequence 245, App
37 16 1.0 391 4 US-09-433-826B-245 Sequence 245, App
38 16 1.0 391 4 US-09-604-287A-245 Sequence 245, App
39 16 1.0 447 4 US-09-795-926-7 Sequence 7, Appli
40 16 1.0 447 4 US-09-795-926-9 Sequence 9, Appli
41 16 1.0 561 4 US-09-795-926-11 Sequence 11, Appl
42 16 1.0 609 4 US-09-252-991A-12054 Sequence 12054, A
43 16 1.0 618 4 US-09-134-001C-210 Sequence 210, App
44 16 1.0 639 4 US-09-252-991A-11921 Sequence 11921, A
45 16 1.0 654 5 PCT-US91-0295A-10 Sequence 10, Appli

ALIGNMENTS

RESULT 1
US-08-943-714-1
; Sequence 1, Application US/08943714
; Patent No. 6187578
; GENERAL INFORMATION:
; APPLICANT: Blinkovsky, Alexander
; APPLICANT: Berk, Randy
; APPLICANT: Rey, Michael
; APPLICANT: Golightly, Elizabeth
; APPLICANT: Klotz, Alan
; APPLICANT: Mathisen, Thomas Erik
; APPLICANT: Dambmann, Claus
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 61875780 No. 6187578disk of No. 6187578th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,714
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4990.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1662 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..1662
; OTHER INFORMATION:

; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 1066
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (218)..(571)
US-09-620-312D-1066

Query Match 1.1%; Score 18; DB 4; Length 716;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1566 CTGGCCCGAGTCAAGAC 1583
|||
DB 65 CTGGCCCGAGTCAAGAC 48

RESULT 3
US-09-328-352-867/c
; Sequence 867, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 867
; LENGTH: 729
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-867

Query Match 1.1%; Score 18; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 711 ATTCTTCATCATTTT 728
|||
DB 328 ATTCTTCATCATTTT 311

RESULT 4
US-09-252-991A-5290/c
; Sequence 5290, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5290
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5290

Query Match 1.0%; Score 17; DB 4; Length 513;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1320 GTTCGGCGGTGAGCCG 1336
|||
DB 310 GTTCGGCGGTGAGCCG 294

RESULT 5
US-09-252-991A-14098
; Sequence 14098, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14098
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14098

Query Match 1.0%; Score 17; DB 4; Length 540;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1289 TCTATGGCGAGCGCGAT 1305
|||
DB 111 TCTATGGCGAGCGCGAT 127

RESULT 6
US-09-252-991A-5185
; Sequence 5185, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5185
; LENGTH: 768
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5185

Query Match 1.0%; Score 17; DB 4; Length 768;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1320 GTTCGGCGGTGAGCCG 1336
|||
DB 312 GTTCGGCGGTGAGCCG 328

RESULT 7
US-09-221-017B-125

```

; Sequence 125, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221.017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 125:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1...1116
; US-09-221-017B-125

Query Match 1.0%; Score 17; DB 4; Length 1116;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 TGTCCACCATCCGGTACA 157
DB 834 TGTCCACCATCCGGTACA 850

RESULT 8
US-08-504-459-7
; Sequence 7, Application US/08504459
; Patent No. 5922563
; GENERAL INFORMATION:
; APPLICANT: Alderete, John F.
; TITLE OF INVENTION: Adhesin Genes and Proteins Involved in
; Trichomonas Vaginalis Cytoadherence
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/504.459
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: UTSK:273/WIM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1214
; US-08-504-459-7

Query Match 1.0%; Score 17; DB 2; Length 1319;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 868 GGATCAAGGCTGTCAA 884
DB 1173 GGATCAAGGCTGTCAA 1189

RESULT 9
US-09-134-001C-373
; Sequence 373, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 373
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-373

Query Match 1.0%; Score 17; DB 4; Length 1347;
Best Local Similarity 100.0%; Pred. No. 53;

```


Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1108 AGTTATTACAAAT 1124
 |||||
 Db 803 AGTTATTACAAAT 819

RESULT 10

US-09-252-991A-13416
 ; Sequence 13416, Application US/09252991A
 ; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 13416

; LENGTH: 1629
 ; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-13416

Query Match 1.0%; Score 17; DB 4; Length 1629;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1585 AATGAACGGCTACAGC 1601
 |||||
 Db 1142 AATGAACGGCTACAGC 1158

RESULT 11

US-09-252-991A-13732/c
 ; Sequence 13732, Application US/09252991A
 ; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 13732

; LENGTH: 1731
 ; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-13732

Query Match 1.0%; Score 17; DB 4; Length 1731;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1585 AATGAACGGCTACAGC 1601
 |||||
 Db 578 AATGAACGGCTACAGC 562

RESULT 12

US-09-399-913-52
 ; Sequence 52, Application US/09399913
 ; Patent No. 6361971

; GENERAL INFORMATION:

; APPLICANT: Rhodes, Kenneth

; APPLICANT: Betty, Maria

; APPLICANT: Ling, Huai-Ping

; APPLICANT: An, Wenqian

; TITLE OF INVENTION: POTASSIUM CHANNEL INTERACTORS AND USES THEREFOR
 ; FILE REFERENCE: MNI-070CP2

; CURRENT APPLICATION NUMBER: US/09/399,913

; CURRENT FILING DATE: 1999-09-21

; EARLIER APPLICATION NUMBER: USN 60/110,277

; EARLIER FILING DATE: 1998-11-30

; EARLIER APPLICATION NUMBER: USN 60/110,033

; EARLIER FILING DATE: 1998-11-25

; EARLIER APPLICATION NUMBER: USN 60/109,333

; EARLIER FILING DATE: 1998-11-20

; EARLIER APPLICATION NUMBER: USN 09/298,731

; EARLIER FILING DATE: 1999-04-23

; EARLIER APPLICATION NUMBER: USN 09/350,614

; EARLIER FILING DATE: 1999-07-09

; EARLIER APPLICATION NUMBER: USN 09/350,874

; EARLIER FILING DATE: 1999-07-09

; NUMBER OF SEQ ID NOS: 73

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 52

; LENGTH: 2051

; TYPE: DNA

; ORGANISM: Rattus sp.

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (85)..(1305)

; US-09-399-913-52

Query Match 1.0%; Score 17; DB 4; Length 2051;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 756 TAGTGTAAATGGTTC 772
 |||||
 Db 1501 TAGTGTAAATGGTTC 1517

RESULT 13

US-09-252-991A-14128/c
 ; Sequence 14128, Application US/09252991A
 ; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 14128

; LENGTH: 2838

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

; US-09-252-991A-14128

Query Match 1.0%; Score 17; DB 4; Length 2838;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1289 TCTATGGCGACCGCAT 1305
 |||||
 Db 1421 TCTATGGCGACCGCAT 1405

RESULT 14

US-09-252-991A-14008
; Sequence 14008, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14008
; LENGTH: 3897
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14008

Query Match 1.0%; Score 17; DB 4; Length 3897;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1289 TCTATGGCGACGCCGAT 1305
|||
Db 2468 TCTATGGCGACGCCGAT 2484
|||

RESULT 15
US-08-309-512-2
; Sequence 2, Application US/08309512
; Patent No. 5759828
; GENERAL INFORMATION:
; APPLICANT: Tal, Ronny
; APPLICANT: Benziman, Moshe
; APPLICANT: Gelfand, David H.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Calhoon, Roger D.
; APPLICANT: Wong, Hing C.
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Pennie & Edmonds
; STREET: 2730 Sand Hill Road
; CITY: Menlo Park
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,512
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,218
; FILING DATE: 29-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bortner, Scott R.
; REGISTRATION NUMBER: 34,298
; REFERENCE/DOCKET NUMBER: 8145-008
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4558 base pairs

US-09-252-991A-14008
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Acetobacter xylinum
US-09-309-512-2

Query Match 1.0%; Score 17; DB 1; Length 4558;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 381 TGATGACTACATCAACC 397
|||
Db 363 TGATGACTACATCAACC 379
|||

Search completed: November 21, 2003, 17:27:35
Job time : 131 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 14:26:12 ; Search time 3271 Seconds
(without alignments)

12393.722 Million cell updates/sec

Title: US-09-712-338-1

Perfect score: 1668

Sequence: 1 atgcgtggtaacgaattctt.....ccagtgtgtatggcatag 1668

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rtd:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	1.3	245	14	CA752967 002000010
2	22	1.3	438	6	AL810229
3	22	1.3	532	14	CB000535
4	22	1.3	561	12	BU235208

5	22	1.3	568	10	BG606737
6	22	1.3	616	12	BJ279570
7	22	1.3	624	14	CA644850
8	22	1.3	681	14	CB000687
9	21	1.3	185	13	BQ377989
10	21	1.3	427	14	CB508631
11	21	1.3	823	29	BZ276557
12	20	1.2	285	10	BB266301
13	20	1.2	319	9	AV427109
14	20	1.2	483	9	AU271072
15	20	1.2	527	28	BZ184493
16	20	1.2	528	13	BQ585524
17	20	1.2	560	12	BQ433922
18	20	1.2	751	29	BZ270192
19	20	1.2	1119	13	BQ072757
20	20	1.2	1137	12	BM810136
21	19	1.1	304	10	BB286068
22	19	1.1	305	9	AV101028
23	19	1.1	313	10	BF290905
24	19	1.1	314	10	BF209547
25	19	1.1	323	10	BF558678
26	19	1.1	324	14	CA543232
27	19	1.1	376	29	CNS007LU
28	19	1.1	381	13	BY406770
29	19	1.1	424	13	BY505128
30	19	1.1	435	12	BJ559889
31	19	1.1	469	28	AZ244328
32	19	1.1	477	10	BG366602
33	19	1.1	486	13	BU802271
34	19	1.1	495	28	BH046325
35	19	1.1	500	9	AA007332
36	19	1.1	502	9	AA797341
37	19	1.1	511	14	CB020176
38	19	1.1	517	28	AO697403
39	19	1.1	524	14	CB020173
40	19	1.1	531	13	BM124273
41	19	1.1	532	28	BH872089
42	19	1.1	536	14	CA734895
43	19	1.1	543	10	BE975750
44	19	1.1	559	13	BW254340
45	19	1.1	580	10	BB609996

ALIGNMENTS

RESULT 1
CA752967
LOCUS
DEFINITION
CA752967
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 245)
Pathan,M.S., Klueva,N., McCartor,K., Zhang,D., Feril,O., Millena,C., Fredricksen,M., Bohnert,H. and Nguyen,H.
Subtracted rice IR62266 leaf drought-stressed cDNA library
Unpublished
Contact: Mark Fredricksen
Department of Plant Biology
University of Illinois
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 2172655473
Email: bohnertlab@life.uiuc.edu.
Location/Qualifiers
L. 245

CA752967 245 bp mRNA linear EST 27-NOV-2002
00200001020.D03_00101218N0.scf IR62266 Oryza sativa cDNA clone
00200001020.D03_00101218N0.scf similar to Hordeum vulgare
carboxypeptidase D precursor, mRNA sequence.

CA752967
GI:25797006

EST

Oryza sativa

Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 245)

Pathan,M.S., Klueva,N., McCartor,K., Zhang,D., Feril,O., Millena,C., Fredricksen,M., Bohnert,H. and Nguyen,H.

Subtracted rice IR62266 leaf drought-stressed cDNA library

Unpublished

Contact: Mark Fredricksen

Department of Plant Biology

University of Illinois

1201 W. Gregory Dr., Urbana, IL 61801, USA

Tel: 2172655473

Email: bohnertlab@life.uiuc.edu.

Location/Qualifiers

L. 245

QY	682	GAGAGCTATGGAGGCACCTATG	703
DB	362	GAGAGCTATGGAGGCACCTATG	383
RESULT 3			
LOCUS	CB000535		
DEFINITION	S345C_E05 Rice cold stress germination cDNA library Oryza sativa (japonica cultivar-group) cDNA 5', mRNA sequence.		
ACCESSION	CB000535		
KEYWORDS	EST.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; Bhrhartidoideae; Oryzeae; Oryza. 1 (bases 1 to 532)		
REFERENCE	de los Reyes,B.G., Morsy,M., Gibbons,J., Varma,T.S.N., Antoine,W., Redus,M., McGrath,J.M. and Malgren,R. Development of a chilling stress EST library of germinating rice (Oryza sativa L. subsp. japonica) enriched with stress-related and novel genes Unpublished		
JOURNAL	Contact: Benildo G. de los Reyes Plant Genomics Lab., Department of Crop, Soil and Environmental Sciences University of Arkansas 115 Plant Science Building, Fayetteville, AR 72701, USA Tel: (479)-575-7465 Fax: (479)-575-8435 Email: breyes@uark.edu Plate: S345C row: E column: 05 Seq primer: T3.		
COMENT	Location/Qualifiers		
FEATURES	1..532		
source	/organism="Oryza sativa (japonica cultivar-group)" /mol_type="mRNA" /db_xref="taxon:39947" /tissue_type="coleotile, radicle, prophyll, immature leaf" /dev_stage="embryo at different stages of germination and shoots of germinated seeds under chilling stress (130C/100C)" /lab_host="SoLR"		
	/note="Organ: seedlings; Vector: Lambda Uni-Zap XR excised as pBluescript; Site 1: EcoRI; Site 2: XhoI; The cDNAs were derived from reverse transcription of mRNA samples from seeds at different stages of germination and seedlings at early phase of growth under chilling stress (130C/100C). The mRNA pool was used as template for double stranded cDNA synthesis using the Stratagene Uni-Zap XR cDNA synthesis and library kit. A total of 150,000 phages were excised from the primary library as pBluescript phagemid clones. Enrichment of the primary excised library with chilling-induced transcripts was performed by hybridizing the primary excised library colony lifts with the PCR-select subtraction product, with cold germinated cDNA as tester and control temperature-germinated cDNA as driver."		
BASE COUNT	155 a 109 C 126 G 142 t		
ORIGIN			
Query Match	1.3%; Score 22; DB 14; Length 532;		
Best Local Similarity	100.0%; Pred.No. 3.4;		
Matches	22; Conservative 0; Mismatches 0; Indels 0; Gaps 0		
QY	682	GAGAGCTATGGAGGCACCTATG	703
DB	114	GAGAGCTATGGAGGCACCTATG	135

```

RESULT 4
BU235208          561 bp  mRNA  linear  EST 05-APR-2002
LOCUS             aestivum cDNA clone whr3j16 5', mRNA sequence.
DEFINITION        Buz235208 Y. Ogiwara unpublished cDNA library, wh_e Triticum
ACCESSION         BU235208
VERSION           BU235208.1 GI:20051723
KEYWORDS          EST.
SOURCE            Triticum aestivum (bread wheat)
ORGANISM          Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 561)
Ogiwara, Y. and Murai, K.
Expressed genes in Triticum aestivum
Unpublished
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..561
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="whr3j16"
                     /tissue_type="seed DPA10"
                     /dev_stage="Feekes' scale 11.2"
                     /clone_lib="Y. Ogiwara unpublished cDNA library, wh_e"
BASE COUNT        139 a 147 c 142 g 133 t
ORIGIN
Query Match       1.3%; Score 22; DB 12; Length 561;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 GAGAGCTATGGAGGGCACTATG 703
Db 216 GAGAGCTATGGAGGGCACTATG 237
|||||
|||||

RESULT 5
BG606737          568 bp  mRNA  linear  EST 17-APR-2001
LOCUS             WHE2452 D07_G14ZS Triticum monococcum early reproductive apex cDNA
DEFINITION        library Triticum monococcum cDNA clone WHE2452_D07_G14, mRNA
ACCESSION         BG606737
VERSION           BG606737.1 GI:13656720
KEYWORDS          EST.
SOURCE            Triticum monococcum
ORGANISM          Triticum monococcum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 568)
Anderson, O.D., Chao, S., Dubcovsky, J., Echenique, V., Han, P.S., Hsia
, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L.,
Stamova, B. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes - Early reproductive apex cDNA library from Triticum
monococcum
Unpublished
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center

```

800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanders@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stragene SK primer.

FEATURES

```

     source          1..568
                     /organism="Triticum monococcum"
                     /mol_type="mRNA"
                     /cultivar="DV92"
                     /db_xref="taxon:4568"
                     /clone="WHE2452_D07_G14"
                     /tissue_type="Early reproductive apex"
                     /dev_stage="Seven week-old plants"
                     /lab_host="E. coli XL0LR"
                     /clone_lib="Triticum monococcum early reproductive apex
                     cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site 1: EcoRI; Site 2: XhoI; The tissue, total RNA, and
                     poly(A) RNA were prepared from apex at double-ridge stage
                     to terminal-spikelet stage during transition from
                     vegetative state to flower state, a cDNA library was made,
                     and the cDNA clones were in vivo excised at the
                     University of California, Davis (V. Echenique, B. Stamova
                     , J. Dubcovsky). Plasmid DNA preparations and DNA
                     sequencing were performed in the OD Anderson lab (all
                     other authors)."
BASE COUNT        102 a 197 c 165 g 104 t
ORIGIN
Query Match       1.3%; Score 22; DB 10; Length 568;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 GAGAGCTATGGAGGGCACTATG 703
Db 458 GAGAGCTATGGAGGGCACTATG 479
|||||
|||||

RESULT 6
BU279570          616 bp  mRNA  linear  EST 09-APR-2002
LOCUS             aestivum cDNA clone whr3j20 5', mRNA sequence.
DEFINITION        Buz279570 Y. Ogiwara unpublished cDNA library, wh_r Triticum
ACCESSION         BU279570
VERSION           BU279570.1 GI:20102057
KEYWORDS          EST.
SOURCE            Triticum aestivum (bread wheat)
ORGANISM          Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 616)
Ogiwara, Y. and Murai, K.
Expressed genes in Triticum aestivum
Unpublished
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..616
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="whr3j20"
                     /tissue_type="root"

```

/dev stage="Peekes' scale 1"
/clone lib="V. Ogihara unpublished cdna library, Wh_r"
145 a 169 c 166 g 136 t

Query Match 1.3%; Score 22; DB 12; Length 616;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 GAGAGCTATGAGGGCACTATG 703
|||||
Db 292 GAGAGCTATGAGGGCACTATG 313
|||||

RESULT 7
LOCUS CA644850 624 bp mRNA linear EST 23-NOV-2002
DEFINITION wrein.pk0093.f11 wrein Triticum aestivum cdna clone
CA644850
VERSION wrein.pk0093.f11 5' end, mRNA sequence.
KEYWORDS CA644850.1 GI:25223146
SOURCE EST.
ORGANISM Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;
Triticeae; Triticum.
1 (bases 1 to 624)
Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
DuPont Wheat cdna Sequence
Unpublished
Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.

FEATURES
source
1..624
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wrein.pk0093.f11"
/tissue_type="root"
/clone_lib="wrein"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: XhoI; Wheat (Triticum aestivum L.) root; normalized from wrein library"

BASE COUNT 161 a 160 c 160 g 136 t 7 others
ORIGIN

Query Match 1.3%; Score 22; DB 14; Length 624;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 GAGAGCTATGAGGGCACTATG 703
|||||
Db 242 GAGAGCTATGAGGGCACTATG 263
|||||

RESULT 8
LOCUS CB000687 681 bp mRNA linear EST 10-JAN-2003
DEFINITION S3450_H12 Rice cold stress germination cdna library Oryza sativa (japonica cultivar-group) cdna 5', mRNA sequence.
CB000687
VERSION CB000687.1 GI:27577992
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 681)
de los Reyes,B.G., Morsy,M., Gibbons,J., Varma,T.S.N., Antoine,W.,
Redus,M., McGrath,J.M. and Halgren,R.
Development of a chilling stress EST library of germinating rice
(Oryza sativa L. subsp. japonica) enriched with stress-related and
novel genes
Unpublished
Contact: Benildo G. de los Reyes
Plant Genomics Lab., Department of Crop, Soil and Environmental
Sciences
University of Arkansas
115 Plant Science Building, Fayetteville, AR 72701, USA
Tel: (479)-575-7465
Fax: (479)-575-8435
Email: breyes@uark.edu
Plate: S345J row: H column: 12
Seq primer: T3.

FEATURES
Location/Qualifiers
1..681
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="PI560247"
/db_xref="taxon:39947"
/tissue_type="coleotile, radicle, prophyll, immature leaf"
/dev_stage="embryo at different stages of germination and
shoots of germinated seeds under chilling stress
(130C/100C)"
/lab_host="SoltR"
/clone_lib="Rice cold stress germination cdna library"
/note="Organ: seedlings; Vector: Lambda Uni-Zap XR excised
as pBluescript; Site_1: EcoRI; Site_2: XhoI; The cDNAs
were derived from reverse transcription of mRNA samples
from seeds at different stages of germination and
seedlings at early phase of growth under chilling stress
(130C/100C). The mRNA pool was used as template for double
stranded cDNA synthesis using the Stratagene Uni-Zap XR
cDNA synthesis and library kit. A total of 150,000 phages
were excised from the primary library as pBluescript
phagemid clones. Enrichment of the primary excised library
with chilling-induced transcripts was performed by
hybridizing the primary excised library colony lifts with
the PCR-select subtraction product, with cold germinated
cDNA as tester and control temperature-germinated cDNA as
driver."

BASE COUNT 193 a 144 c 158 g 186 t
ORIGIN

Query Match 1.3%; Score 22; DB 14; Length 681;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 GAGAGCTATGAGGGCACTATG 703
|||||
Db 114 GAGAGCTATGAGGGCACTATG 135
|||||

RESULT 9
LOCUS BQ377989/c 185 bp mRNA linear EST 21-MAY-2002
DEFINITION MRL-UM0009-010600-018-e11 UM0009 Homo sapiens cdna, mRNA sequence.
BQ377989
VERSION BQ377989.1 GI:21053503
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 185)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

/dev stage="Peekes' scale 1"
/clone lib="V. Ogihara unpublished cdna library, Wh_r"
145 a 169 c 166 g 136 t

Query Match 1.3%; Score 22; DB 12; Length 616;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 GAGAGCTATGAGGGCACTATG 703
|||||
Db 292 GAGAGCTATGAGGGCACTATG 313
|||||

RESULT 7
LOCUS CA644850 624 bp mRNA linear EST 23-NOV-2002
DEFINITION wrein.pk0093.f11 wrein Triticum aestivum cdna clone
CA644850
VERSION wrein.pk0093.f11 5' end, mRNA sequence.
KEYWORDS CA644850.1 GI:25223146
SOURCE EST.
ORGANISM Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;
Triticeae; Triticum.
1 (bases 1 to 624)
Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
DuPont Wheat cdna Sequence
Unpublished
Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.

FEATURES
source
1..624
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wrein.pk0093.f11"
/tissue_type="root"
/clone_lib="wrein"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: XhoI; Wheat (Triticum aestivum L.) root; normalized from wrein library"

BASE COUNT 161 a 160 c 160 g 136 t 7 others
ORIGIN

Query Match 1.3%; Score 22; DB 14; Length 624;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 GAGAGCTATGAGGGCACTATG 703
|||||
Db 242 GAGAGCTATGAGGGCACTATG 263
|||||

RESULT 8
LOCUS CB000687 681 bp mRNA linear EST 10-JAN-2003
DEFINITION S3450_H12 Rice cold stress germination cdna library Oryza sativa (japonica cultivar-group) cdna 5', mRNA sequence.
CB000687
VERSION CB000687.1 GI:27577992
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR1&t2=MR1-UM0009-010600-018-ell&t3=2000-06-01&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 18
High quality sequence stop: 185.

FEATURES

source
Location/Qualifiers
1..185
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="UM0009"

/note="Organ: uterus; Vector: puc18; Site 1: Smal; Site 2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 37 a 63 c 51 g 34 t

ORIGIN

Query Match 1.3%; Score 21; DB 13; Length 185;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1076 TTGGGCATCCATATGATGACC 1096
|||||

Db 74 TTGGGCATCCATATGATGACC 54

RESULT 10 CB508631/c 427 bp mRNA linear EST 16-MAY-2003
LOCUS ssalob502032 reproductive Salmo salar cDNA, mRNA sequence.
DEFINITION CB508631

ACCESSION CB508631.1 GI:29320357

VERSION EST.

KEYWORDS Salmo salar (Atlantic salmon)

SOURCE Salmo salar

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.

REFERENCE 1 (bases 1 to 427)

AUTHORS GRASP Consortium, Davidson, W.S., Koop, B.F. and <http://web.uvic.ca/cbr/grasp>.

TITLE A survey of Salmo salar transcripts from high complexity cDNA libraries

JOURNAL Unpublished

COMMENT Contact: Koop BF
Centre for Biomedical Research
University of Victoria
PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
Tel: 250 472 4067

Fax: 250 472 4075
Email: dkoop@uvic.ca

Genome Sciences Centre, BC Cancer Agency cDNA preparation, sequencing and bioinformatics: Y Butterfield, R Kirkpatrick, J Asano, N Girn, R Guin, D Lee, S Lee, T Olson, P Pandoh, A Prabhu, D Smalhus, L Spence, J Stott, S Taylor, G Yang, J Schein, S Jones and M Marra.

POLYA=Yes.

FEATURES

source
Location/Qualifiers
1..427
/organism="Salmo salar"
/mol_type="mRNA"
/strain="McConnell"
/db_xref="taxon:8030"
/clone_lib="reproductive"
/note="Vector: pBlueScriptIIISK+; Library Creator: Kristian R von Schalburg; Atlantic salmon tissue contributors: Carlo Biagi, Mitch Uh and Robert Devlin (DFO Vancouver, B.C.), Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton, B.C.), Rachel Roper (University of Victoria)"

BASE COUNT 75 a 130 c 109 g 113 t

ORIGIN

Query Match 1.3%; Score 21; DB 14; Length 427;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1157 GGGTCAACATCACTACACCC 1177
|||||

Db 197 GGGTCAACATCACTACACCC 177

RESULT 11

BZ276557/c

LOCUS

DEFINITION

CH230-294N6.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone

CH230-294N6, genomic survey sequence.

ACCESSION BZ276557

VERSION BZ276557.1 GI:23999653

KEYWORDS GSS.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 823)

AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P., and Fraser, C.M.

TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment

JOURNAL Unpublished

COMMENT Other GSSs: CH230-294N6.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 294 row: N column: 6
Seq primer: T7
Class: BAC ends.

FEATURES

source
Location/Qualifiers
1..823
/organism="Rattus norvegicus"
/mol_type="genomic DNA"

```

/strain="BN/SSNHsd/MCW"
/db xref="taxon:10116"
/clone="CH230-294N6"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
Pieter de Jong"
BASE COUNT      279 a   174 c   200 g   170 t
ORIGIN
Query Match      1.3%; Score 21; DB 29; Length 823;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 ATTTCCTCAGTGTACCCCTT 35
|||||
Db 29 ATTTCCTCAGTGTACCCCTT 9

RESULT 12
BE266301 285 bp mRNA linear EST 07-JUL-2000
LOCUS BE266301 RIKEN full-length enriched, 10 days neonate cortex Mus
DEFINITION musculus cDNA clone A830025A06 3' similar to AF134238 Mus musculus
PL6 protein (p16) mRNA, mRNA sequence.
ACCESSION BE266301
VERSION BE266301.1 GI:8962763
KEYWORDS Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 285)
Konno.H., Aizawa.K., Akahira.S., Akiyama.J., Arakawa.T., Carninci
P., Endo.T., Fukuda.S., Fukunishi.Y., Hara.A., Hayatsu.N.,
Hirozane.T., Hori.F., Ishii.Y., Ishikawa.J., Ishikawa.T., Itoh.M.,
Izawa.M., Kadota.K., Kagawa.I., Kai.C., Kawai.J., Kikuchi.N.,
Kiyosawa.H., Kojima.Y., Kondo.S., Koya.S., Kurihara.C., Kusakabe.M.,
Matsuyama.T., Miki.R., Mizuno.Y., Nakamura.M., Oda.H., Okazaki.Y.,
Ono.T., Owa.C., Saito.H., Sakai.C., Sato.K., Shibata.K., Shibata
Y., Shigemoto.Y., Shinagawa.A., Shiraki.T., Sogabe.Y., Sugahara.Y.,
Suzuki.H., Suzuki.H., Tagawa.A., Takahashi.F., Tominaga.N., Toya
T., Teunoda.Y., Watahiki.A., Watanabe.S., Yamamura.T., Yamanaka.I.,
Yano.R., Yasunishi.A., Yokota.T., Yoshida.K., Yoshiki.A., Yoshino
M., Muramatsu.M. and Hayashizaki.Y.
RIKEN Mouse ESTs (Konno.H., et al.)
Unpublished
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermosensitization and thermostabilization of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru.Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
Y., and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

Location/Qualifiers
1. .285
/organism="Mus musculus"
/mol_type="mRNA"
/db xref="taxon:10090"
/clone="A830025A06"
/tissue_type="cortex"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 10 days neonate
cortex"
/note="Site 1: Sall; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCAGAGCTCTTTTITTTTIVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTTAATTAATTAATCCGCCGCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."
BASE COUNT      64 a   87 c   56 g   78 t
ORIGIN
Query Match      1.2%; Score 20; DB 10; Length 285;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 CTATTCCTGTCCAGCCATT 446
|||||
Db 3 CTATTCCTGTCCAGCCATT 22

RESULT 13
AV427109/c
LOCUS AV427109 Lotus japonicus young plants (two-week old) Lotus
DEFINITION japonicus cDNA clone MWM076c05_r 5', mRNA sequence.
ACCESSION AV427109
VERSION AV427109.1 GI:7786723
KEYWORDS EST.
SOURCE Lotus japonicus
ORGANISM Lotus japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids 1; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
1 (bases 1 to 319)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
Generation of 7137 non-redundant expressed sequence tags from a
legume, Lotus japonicus
DNA Res. 7 (2), 127-130 (2000)
MEDLINE 20277479
PUBMED 10819328
COMMENT Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakam@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers
1. .319
/organism="Lotus japonicus"
/mol_type="mRNA"
/db_xref="taxon:34305"
/clone="MWM076c05_r"
/dev_stage="young plants (two-week old)"
/clone_lib="Lotus japonicus young plants (two-week old)"

```



```

/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI; isolate=Miyakojima MG-20"
BASE COUNT      87 a      69 c      74 g      89 t
ORIGIN
Query Match      1.2%; Score 20; DB 9; Length 319;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1013 ACATGTCAGGAGCAATGTT 1032
Db 131 ACATGTCAGGAGCAATGTT 112

RESULT 14
AU271072/c
LOCUS      AU271072
DEFINITION AU271072 VS Dictyostelium discoideum cDNA clone VSK365 5', mRNA
sequence.
ACCESSION AU271072
VERSION    AU271072.1 GI:20529870
KEYWORDS   EST.
SOURCE     Dictyostelium discoideum
ORGANISM   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

REFERENCE 1 (bases 1 to 483)
AUTHORS   Urushihara,H., Morio,T., Saito,T., Kotiki,E., Ochiai,H., Maeda,M.,
          Takeuchi,I., Kohara,Y. and Tanaka,Y.
TITLE     Population analysis of cDNAs from unicellular and multicellular
          stages of Dictyostelium discoideum
JOURNAL    Unpublished
COMMENT    Contact: Hideko Urushihara
          Institute of Biological Sciences
          University of Tsukuba
          1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
          Tel: 81-298-53-4664
          Fax: 81-298-53-6614
          Email: hideko@biol.tsukuba.ac.jp.

FEATURES
source
1..483
Location/Qualifiers
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="VSK365"
/sex="mat A"
/dev_stage="vegetative"
/clone_lib="VS"

BASE COUNT      180 a      74 c      52 g      175 t      2 others
ORIGIN
Query Match      1.2%; Score 20; DB 9; Length 483;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 TTACGACGAGCAATGAGAGA 744
Db 100 TTACGACGAGCAATGAGAGA 81

RESULT 15
BZ184493
LOCUS      BZ184493
DEFINITION CH230-485P4.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
sequence.
ACCESSION BZ184493
VERSION    BZ184493.1 GI:23834121
KEYWORDS   GSS.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.

```

REFERENCE

1 (bases 1 to 527)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,
A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.

TITLE
JOURNAL
COMMENT

Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished
Other_GSSs: CH230-485P4.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
plate: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 485 row: P column: 4
Seq primer: T7
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1..527
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-485P4"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
Pieter de Jong"

BASE COUNT 162 a 103 c 112 g 150 t
ORIGIN

Query Match 1.2%; Score 20; DB 28; Length 527;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 AGCCAGACATACCCAGAAA 283
Db 116 AGCCAGACATACCCAGAAA 135

Search completed: November 21, 2003, 17:25:15
Job time : 3275 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 14:38:17 ; Search time 577 Seconds
(without alignments)
9447.773 Million cell updates/sec

Title: US-09-712-338-1
Perfect score: 1668
Sequence: 1 atgcgtggtagaattctt.....ccagtgtgtatggcatag 1668

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 2169961 seqs, 1634102185 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	1.3	2799	12	US-10-241-009-18
2	21	1.3	2799	12	US-10-241-009-53
3	21	1.3	2799	12	US-10-190-434B-18
4	21	1.3	2799	12	US-10-190-434B-53
5	21	1.3	2799	12	US-10-190-305A-11
6	21	1.3	2799	12	US-10-190-305A-74
7	21	1.3	3205	12	US-10-241-009-17
8	21	1.3	3205	12	US-10-190-434B-17
9	21	1.3	3496	12	US-10-241-009-15
10	21	1.3	3496	12	US-10-190-434B-15
11	21	1.3	3999	12	US-10-241-009-9
12	21	1.3	3999	12	US-10-241-009-10
13	21	1.3	3999	12	US-10-241-009-11
14	21	1.3	3999	12	US-10-190-434B-9
15	21	1.3	3999	12	US-10-190-434B-10
16	21	1.3	3999	12	US-10-190-434B-11

17	21	1.3	4773	12	US-10-241-009-16	Sequence 16, Appl
18	21	1.3	4773	12	US-10-241-009-55	Sequence 55, Appl
19	21	1.3	4773	12	US-10-190-434B-16	Sequence 16, Appl
20	21	1.3	4773	12	US-10-190-434B-55	Sequence 55, Appl
21	21	1.3	4773	12	US-10-190-305A-10	Sequence 10, Appl
22	21	1.3	4773	12	US-10-190-305A-76	Sequence 76, Appl
23	21	1.3	5274	12	US-10-241-009-12	Sequence 12, Appl
24	21	1.3	5274	12	US-10-190-434B-12	Sequence 9, Appl
25	21	1.3	5274	12	US-10-190-305A-9	Sequence 54, Appl
26	21	1.3	5283	12	US-10-241-009-54	Sequence 54, Appl
27	21	1.3	5283	12	US-10-190-305A-75	Sequence 75, Appl
28	21	1.3	5283	12	US-10-190-305A-75	Sequence 874, App
29	19	1.1	594	14	US-10-156-761-874	Sequence 1242, Ap
30	19	1.1	989	10	US-09-974-300-1242	Sequence 25, Appl
31	19	1.1	3563	10	US-09-900-237-25	Sequence 2883, Ap
32	19	1.1	14055	14	US-10-156-761-2883	Sequence 15102, A
33	19	1.1	125746	14	US-10-156-761-15102	Sequence 1, Appl
34	19	1.1	9025608	14	US-10-156-761-1	Sequence 11878, A
35	19	1.1	9025608	14	US-09-878-574-11878	Sequence 3692, Ap
36	18	1.1	299	10	US-09-796-692-3692	Sequence 3692, Ap
37	18	1.1	408	10	US-10-040-862-3692	Sequence 1066, Ap
38	18	1.1	408	14	US-10-037-270-1066	Sequence 261103, A
39	18	1.1	716	14	US-10-027-632-261103	Sequence 261104, A
40	18	1.1	1977	12	US-10-027-632-261103	Sequence 261104, A
41	18	1.1	1977	12	US-10-027-632-261103	Sequence 261104, A
42	18	1.1	1977	13	US-10-027-632-261104	Sequence 159, Appl
43	18	1.1	1977	13	US-10-027-632-261104	Sequence 22, Appl
44	18	1.1	3128	12	US-10-240-965-159	
45	18	1.1	3380	14	US-10-128-714-22	

ALIGNMENTS

RESULT 1

US-10-241-009-18
; Sequence 18, Application US/10241009
; Publication No. US20030170614A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 2799
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagTatRevNef.opt_B
US-10-241-009-18

Query Match 1.3%; Score 21; DB 12; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1561 AAGATCTGGCCCACTACAAG 1581
|||||
Db 1318 AAGATCTGGCCCACTACAAG 1338

RESULT 2

US-10-241-009-53
; Sequence 53, Application US/10241009
; Publication No. US20030170614A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan

```
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 2799
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefGag B
US-10-190-434B-53

Query Match          1.3%; Score 21; DB 12; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
    |||||
Db 2602 AAGATCTGGCCCGAGCTACAAG 2622

RESULT 5
US-10-190-305A-11
; Sequence 11, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2799
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagTatRevNef.opt_B
US-10-190-305A-11

Query Match          1.3%; Score 21; DB 12; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
    |||||
Db 1318 AAGATCTGGCCCGAGCTACAAG 1338

RESULT 6
US-10-190-305A-74
; Sequence 74, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 74
; LENGTH: 2799
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefGag B
US-10-190-305A-74
```

```
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 2799
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefGag B
US-10-241-009-53

Query Match          1.3%; Score 21; DB 12; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
    |||||
Db 2602 AAGATCTGGCCCGAGCTACAAG 2622

RESULT 3
US-10-190-434B-18
; Sequence 18, Application US/10190434B
; Publication No. US20030194800A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.20
; CURRENT APPLICATION NUMBER: US/10/190,434B
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 2799
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagTatRevNef.opt_B
US-10-190-434B-18

Query Match          1.3%; Score 21; DB 12; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
    |||||
Db 1318 AAGATCTGGCCCGAGCTACAAG 1338

RESULT 4
US-10-190-434B-53
; Sequence 53, Application US/10190434B
; Publication No. US20030194800A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.20
; CURRENT APPLICATION NUMBER: US/10/190,434B
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 68
```

Query Match 1.3%; Score 21; DB 12; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCAGCTACAAG 1581
DB 2602 AAGATCTGGCCAGCTACAAG 2622

RESULT 7

US-10-241-009-17
; Sequence 17, Application US/10241009
; Publication No. US20030170614A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 3205
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagRTmut.SF2
US-10-241-009-17

Query Match 1.3%; Score 21; DB 12; Length 3205;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCAGCTACAAG 1581
DB 1324 AAGATCTGGCCAGCTACAAG 1344

RESULT 8

US-10-190-434B-17
; Sequence 17, Application US/10190434B
; Publication No. US20030194800A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; FILE REFERENCE: 2300-1621.20
; CURRENT APPLICATION NUMBER: US/10/190,434B
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 3205
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagRTmut.SF2
US-10-190-434B-17

Query Match 1.3%; Score 21; DB 12; Length 3205;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCAGCTACAAG 1581
DB 1324 AAGATCTGGCCAGCTACAAG 1344

RESULT 9

US-10-241-009-15
; Sequence 15, Application US/10241009
; Publication No. US20030170614A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 3496
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagProtInaRTmut.SF2
US-10-241-009-15

Query Match 1.3%; Score 21; DB 12; Length 3496;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCAGCTACAAG 1581
DB 1318 AAGATCTGGCCAGCTACAAG 1338

RESULT 10

US-10-190-434B-15
; Sequence 15, Application US/10190434B
; Publication No. US20030194800A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; FILE REFERENCE: 2300-1621.20
; CURRENT APPLICATION NUMBER: US/10/190,434B
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 3496
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagProtInaRTmut.SF2
US-10-190-434B-15

Query Match 1.3%; Score 21; DB 12; Length 3496;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCAGCTACAAG 1581
DB 1318 AAGATCTGGCCAGCTACAAG 1338

RESULT 11

US-10-241-009-9
; Sequence 9, Application US/10241009
; Publication No. US20030170614A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan

```
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3999
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagCompPolmutAtt.SF2
US-10-241-009-9

Query Match          1.3%; Score 21; DB 12; Length 3999;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
    |||||
Db 1324 AAGATCTGGCCCGAGCTACAAG 1344

RESULT 12
US-10-241-009-10
; Sequence 10, Application US/10241009
; Publication No. US20030170614A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 3999
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagCompPolmutAtt.SF2
US-10-241-009-10

Query Match          1.3%; Score 21; DB 12; Length 3999;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
    |||||
Db 1324 AAGATCTGGCCCGAGCTACAAG 1344

RESULT 13
US-10-241-009-11
; Sequence 11, Application US/10241009
; Publication No. US20030170614A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
```

```
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 3999
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagCompPolmutIna.SF2
US-10-241-009-11

Query Match          1.3%; Score 21; DB 12; Length 3999;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
    |||||
Db 1324 AAGATCTGGCCCGAGCTACAAG 1344

RESULT 14
US-10-190-434B-9
; Sequence 9, Application US/10190434B
; Publication No. US20030194800A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.20
; CURRENT APPLICATION NUMBER: US/10/190,434B
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3999
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagCompPolmut.SF2
US-10-190-434B-9

Query Match          1.3%; Score 21; DB 12; Length 3999;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCGAGCTACAAG 1581
    |||||
Db 1324 AAGATCTGGCCCGAGCTACAAG 1344

RESULT 15
US-10-190-434B-10
; Sequence 10, Application US/10190434B
; Publication No. US20030194800A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.20
; CURRENT APPLICATION NUMBER: US/10/190,434B
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 3999
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

Query Match 1.3%; Score 21; DB 12; Length 3999;
Best Local Similarity 100.0%; Pred.No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 AAGATCTGGCCCAGCTACAAG 1581
|||
D_b 1324 AAGATCTGGCCCAGCTACAAG 1344

Search completed: November 21, 2003, 17:37:28
Job time : 587 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_r2p model

Run on: November 21, 2003, 17:51:05 ; Search time 55 Seconds
(without alignments)

11073.063 Million cell updates/sec

Title: US-09-712-338-1

Perfect score: 555

Sequence: 1 atgcgtggtacgaattctt.....ccagtggtgtatggaatg 1668

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 666188 seqs, 182559486 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1261060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_r2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09712338/runat_17112003_170227_10843/app_query.fasta_1.1863
-DB=Published Applications AA -QPMT=fastan -SUFFIX=oli.rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-ALIGN=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1
-MAXLEN=2000000000 -USER=US09712338 @CGN 1.1.15 @runat_17112003_170227_10843
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description

1	8	1.4	91	9	US-09-815-242-5914	Sequence 5914, App
2	8	1.4	91	9	US-09-815-242-12664	Sequence 12664, A
3	8	1.4	91	9	US-09-815-242-12964	Sequence 12964, A
4	8	1.4	91	9	US-09-815-242-13160	Sequence 13160, A
5	8	1.4	102	10	US-09-989-920-242	Sequence 242, App
6	8	1.4	153	9	US-09-864-761-39044	Sequence 39044, A
7	8	1.4	209	12	US-10-294-039-4	Sequence 4, Appli
8	8	1.4	243	12	US-10-259-165-116	Sequence 116, App
9	8	1.4	243	15	US-10-167-015-10	Sequence 10, Appl
10	8	1.4	334	15	US-10-156-761-10653	Sequence 10653, A
11	8	1.4	356	15	US-10-085-233B-4	Sequence 4, Appli
12	8	1.4	406	12	US-10-194-919-35	Sequence 35, Appl
13	8	1.4	411	9	US-09-901-252-16	Sequence 16, Appl
14	8	1.4	421	9	US-09-420-785A-4	Sequence 4, Appli
15	8	1.4	421	9	US-09-901-252-15	Sequence 15, Appl
16	8	1.4	449	10	US-09-768-781-3	Sequence 3, Appli
17	8	1.4	449	10	US-09-768-781-7	Sequence 7, Appli
18	8	1.4	462	10	US-09-768-781-4	Sequence 4, Appli
19	8	1.4	480	10	US-09-736-457-336	Sequence 336, App
20	8	1.4	480	10	US-09-902-941-336	Sequence 336, App
21	8	1.4	480	10	US-09-849-626-336	Sequence 336, App
22	8	1.4	480	11	US-09-476-300-336	Sequence 336, App
23	8	1.4	480	12	US-10-113-872-336	Sequence 336, App
24	8	1.4	480	14	US-10-084-018-8	Sequence 8, Appli
25	8	1.4	480	15	US-10-017-754-336	Sequence 336, App
26	8	1.4	492	9	US-09-925-302-774	Sequence 774, App
27	8	1.4	492	11	US-09-796-753-42	Sequence 42, Appl
28	8	1.4	605	10	US-09-801-368-428	Sequence 428, App
29	8	1.5	760	15	US-10-231-778-224	Sequence 224, App
30	8	1.4	2991	12	US-09-970-944-22	Sequence 76, Appl
31	8	1.4	4349	12	US-10-174-677-76	Sequence 4, Appli
32	8	1.4	4349	12	US-09-970-944-4	Sequence 18, Appl
33	8	1.4	4349	12	US-09-970-944-18	Sequence 19, Appl
34	8	1.4	4349	12	US-09-970-944-19	Sequence 15, Appl
35	8	1.4	4349	15	US-10-160-758-15	Sequence 96, Appl
36	7	1.3	16	12	US-10-158-825-96	Sequence 96, Appl
37	7	1.3	16	15	US-10-158-847-96	Sequence 131, App
38	7	1.3	17	15	US-10-062-710-131	Sequence 9, Appli
39	7	1.3	24	12	US-10-231-894-9	Sequence 9, Appli
40	7	1.3	24	12	US-10-231-889-9	Sequence 39195, A
41	7	1.3	27	9	US-09-864-761-39195	Sequence 24, Appl
42	7	1.3	34	9	US-09-776-490-24	Sequence 34863, A
43	7	1.3	34	9	US-09-776-491-24	Sequence 180, App
44	7	1.3	43	9	US-09-864-761-34863	
45	7	1.3	44	11	US-09-948-783-180	

ALIGNMENTS

RESULT 1

US-09-815-242-5914
; Sequence 5914, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 60/207,727
; PRIOR FILING DATE: 2000-05-26

Alignment Scores:	156	91
Pred. No.:	8	8
Matches:	8	8

APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13160
LENGTH: 91
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-13160

Alignment Scores:
Pred. No.: 156 Length: 91
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 9 Gaps: 0

US-09-712-338-1 (1-1668) x US-09-815-242-13160 (1-91)
QY 767 GGTTCAGCTTAATTCACCTC 790
Db 69 ValPheSerLeuSerThrLeu 76

RESULT 5
US-09-989-920-242
Sequence 242, Application US/09989920
Patent No. US20020172957A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Herve
APPLICANT: Chen, Sei-Yu
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro
FILE REFERENCE: DEX-0291
CURRENT APPLICATION NUMBER: US/09/989,920
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/252,500
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 284
SOFTWARE: PatentIn version 3.1
SEQ ID NO 242
LENGTH: 102
TYPE: PRT
ORGANISM: Homo sapien
US-09-989-920-242

Alignment Scores:
Pred. No.: 154 Length: 102
Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 10 Gaps: 0

US-09-712-338-1 (1-1668) x US-09-989-920-242 (1-102)
QY 1256 AGATCTTCTCTCCCGTGGCTG 1279
Db 7 ArgSerLeuLeuSerProCysVal 14

RESULT 6
US-09-864-761-39044
Sequence 39044, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 39044
LENGTH: 153
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005950.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
 OTHER INFORMATION: EST HUMAN HIT: BE938266.1, EVALUE 1.70e-02
 OTHER INFORMATION: SWISSPROT HIT: O95180, EVALUE 7.60e+00
 US-09-864-761-39044

Alignment Scores:
 Pred. No.: 147 Length: 153
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 9 Gaps: 0

US-09-712-338-1 (1-1668) x US-09-864-761-39044 (1-153)

QY 233 CCGAGTCCCATACCTTCTCTGGT 256
 |||||
 Db 3 ProSerProIleProSerSerGly 10

RESULT 7

US-10-294-039-4
 ; Sequence 4, Application US/10294039
 ; Publication No. US20030134814A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Carotoll, Joseph M.
 ; TITLE OF INVENTION: METHOD OF USING 18080, A HUMAN SERINE
 ; TITLE OF INVENTION: CARBOXYPEPTIDE FAMILY MEMBER
 ; FILE REFERENCE: MP101-230PIRM
 ; CURRENT APPLICATION NUMBER: US/10/294,039
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: 60/338,587
 ; PRIOR FILING DATE: 2001-11-13
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 209
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(209)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-10-294-039-4

Alignment Scores:
 Pred. No.: 141 Length: 209
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 12 Gaps: 0

US-09-712-338-1 (1-1668) x US-10-294-039-4 (1-209)

QY 298 TTGTGTTCAATGGTGGCCCTGGA 321
 |||||
 Db 33 LeuTrpLeuAsnGlyGlyProgly 40

RESULT 8

US-10-259-165-116
 ; Sequence 116, Application US/10259165
 ; Publication No. US20030135888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Zhu, Tong
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Chang, Hur-song
 ; APPLICANT: Briggs, Steven P.
 ; APPLICANT: Cooper, Bret
 ; APPLICANT: Glazebrook, Jane
 ; APPLICANT: Goff, Stephen A.

APPLICANT: Katagiri, Fumiyaki
 APPLICANT: Kreps, Joel
 APPLICANT: Moughamer, Todd
 APPLICANT: Provart, Nicholas
 APPLICANT: Ricke, Darrell
 TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
 FILE REFERENCE: 70030-NP
 CURRENT APPLICATION NUMBER: US/10/259,165
 CURRENT FILING DATE: 2002-09-26
 PRIOR APPLICATION NUMBER: US 60/370,620
 PRIOR FILING DATE: 2002-04-04
 PRIOR APPLICATION NUMBER: US 60/368,327
 PRIOR FILING DATE: 2002-03-27
 PRIOR APPLICATION NUMBER: US 60/325,277
 PRIOR FILING DATE: 2001-09-26
 NUMBER OF SEQ ID NOS: 782
 SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
 SEQ ID NO 116
 LENGTH: 243
 TYPE: PRT
 ORGANISM: Oryza sativa
 FEATURE:
 NAME/KEY: X region
 LOCATION: (113)...(114)
 OTHER INFORMATION: Xaa = any naturally occurring amino acid
 FEATURE:
 NAME/KEY: X region
 LOCATION: (121)...(121)
 OTHER INFORMATION: Xaa = any naturally occurring amino acid
 FEATURE:
 NAME/KEY: X region
 LOCATION: (127)...(127)
 OTHER INFORMATION: Xaa = any naturally occurring amino acid
 FEATURE:
 NAME/KEY: X region
 LOCATION: (225)...(225)
 OTHER INFORMATION: Xaa = any naturally occurring amino acid
 US-10-259-165-116

Alignment Scores:
 Pred. No.: 139 Length: 243
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 12 Gaps: 0

US-09-712-338-1 (1-1668) x US-10-259-165-116 (1-243)

QY 176 CGCTCTGGAGACTACCCGGGTG 199
 |||||
 Db 151 AlaserAlaArgLeuProArgVal 158

RESULT 9

US-10-167-015-10
 ; Sequence 10, Application US/10167015
 ; Publication No. US20030056249A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Simmons, Carl R.
 ; APPLICANT: Gordon-Kamm, William J.
 ; APPLICANT: Johal, Gurmukh
 ; APPLICANT: Acevedo, Pedro A. Navarro
 ; APPLICANT: Tao, Yumin
 ; TITLE OF INVENTION: Anti-Apoptosis Genes and Methods of Use
 ; TITLE OF INVENTION: Thereof
 ; FILE REFERENCE: 1388
 ; CURRENT APPLICATION NUMBER: US/10/167,015
 ; CURRENT FILING DATE: 2002-06-11
 ; PRIOR APPLICATION NUMBER: US 60/297,478
 ; PRIOR FILING DATE: 2001-06-12
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 10

```

; LENGTH: 243
; TYPE: PRT
; ORGANISM: zea mays
US-10-167-015-10

Alignment Scores:
Pred. No.: 139          Length: 243
Score: 8.00           Matches: 8
Percent Similarity: 100.00%  Conservatives: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.44%          Indels: 0
DB: 15                Gaps: 0

US-09-712-338-1 (1-1668) x US-10-167-015-10 (1-243)

QY 758 GTGTTAATGTTGTCAGCTTAATT 781
Db 167 ValLeuMetValPheSerLeulle 174

RESULT 10
US-10-156-761-10653
; Sequence 10653, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10653
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10653

Alignment Scores:
Pred. No.: 134          Length: 334
Score: 8.00           Matches: 8
Percent Similarity: 100.00%  Conservatives: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.44%          Indels: 0
DB: 15                Gaps: 0

US-09-712-338-1 (1-1668) x US-10-156-761-10653 (1-334)

QY 1392 GCCCTGAAGTCAACGGCGTGA 1415
Db 224 AlaProGluSerGlnArgArg 231

RESULT 11
US-10-085-233B-4
; Sequence 4, Application US/10085233B
; Publication No. US20030087249A1
; GENERAL INFORMATION:
; APPLICANT: GLUCKSMANN, MARIA ALEXANDRA
; TITLE OF INVENTION: 93870, A HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTOR AND USES THEREFOR
; FILE REFERENCE: MPI2001-021PIRCF1M
; CURRENT APPLICATION NUMBER: US/10/085,233B
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/272,677
; PRIOR FILING DATE: 2001-03-01

```

```

; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Mus muscalis
US-10-085-233B-4

Alignment Scores:
Pred. No.: 133          Length: 356
Score: 8.00           Matches: 8
Percent Similarity: 100.00%  Conservatives: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.44%          Indels: 0
DB: 15                Gaps: 0

US-09-712-338-1 (1-1668) x US-10-085-233B-4 (1-356)

QY 170 CAGAGGGCGTCTGGAGACTACCC 193
Db 338 GlnArgAlaSerAlaArgLeuPro 345

RESULT 12
US-10-194-919-35
; Sequence 35, Application US/10194919
; Publication No. US20030145350A1
; GENERAL INFORMATION:
; APPLICANT: GVS Gesellschaft fuer Erwerb und Verwertung von Sc
; APPLICANT: Spener, Friedrich
; APPLICANT: Abbadi, Amine
; APPLICANT: Brummel, Monika
; TITLE OF INVENTION: METHOD FOR INCREASING THE CONTENT OF FATTY ACIDS IN PLANTS AND
; FILE REFERENCE: 1617.018US1
; CURRENT APPLICATION NUMBER: US/10/194,919
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: PCT/EP01/00289
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: DE 100 00 978.6
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Brassica napus
US-10-194-919-35

Alignment Scores:
Pred. No.: 131          Length: 406
Score: 8.00           Matches: 8
Percent Similarity: 100.00%  Conservatives: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.44%          Indels: 0
DB: 12                Gaps: 0

US-09-712-338-1 (1-1668) x US-10-194-919-35 (1-406)

QY 577 ACCAATCTTGCCGACGAGCGCGCT 600
Db 121 ThrAsnLeuAlaAlaGluAlaAla 128

RESULT 13
US-09-901-252-16
; Sequence 16, Application US/09901252
; Patent No. US2002002658A1
; GENERAL INFORMATION:
; APPLICANT: Chapple, Clint
; TITLE OF INVENTION: Genes Encoding Sinapoylglucose:Malate Sinapoyltransferase and Meth
; TITLE OF INVENTION: Use
; FILE REFERENCE: N1422-004
; CURRENT APPLICATION NUMBER: US/09/901,252
; CURRENT FILING DATE: 2001-07-09

```

; PRIOR APPLICATION NUMBER: US 60/216593
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Saccharomyces
US-09-901-252-15
Alignment Scores:
Pred. No.: 130 Length: 421
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 9 Gaps: 0
US-09-712-338-1 (1-1668) x US-09-901-252-15 (1-421)
QY 298 TTGTGGTTGAATGGTGGCCCTGGA 321
Db 48 LeutrpleuAsnGlyProGly 55
Search completed: November 21, 2003, 17:58:41
Job time : 59 secs

; PRIOR APPLICATION NUMBER: US 60/216593
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Triticales
US-09-901-252-16
Alignment Scores:
Pred. No.: 131 Length: 411
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 9 Gaps: 0
US-09-712-338-1 (1-1668) x US-09-901-252-16 (1-411)
QY 298 TTGTGGTTGAATGGTGGCCCTGGA 321
Db 57 LeutrpleuAsnGlyProGly 64
RESULT 14
US-09-420-785A-4
; Sequence 4, Application US/09420785A
; Patent No. US20010010923A1
; GENERAL INFORMATION:
; APPLICANT: MORTENSEN, UFFE
; APPLICANT: OLESEN, KJELD
; APPLICANT: STENNICKE, HENNING
; APPLICANT: SORESENSEN, STEEN B.
; APPLICANT: BREDDAM, KLAUS
; TITLE OF INVENTION: MODIFIED CARBOXYPEPTIDASE
; FILE REFERENCE: 089187/0109
; CURRENT APPLICATION NUMBER: US/09/420,785A
; CURRENT FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-420-785A-4
Alignment Scores:
Pred. No.: 130 Length: 421
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 9 Gaps: 0
US-09-712-338-1 (1-1668) x US-09-420-785A-4 (1-421)
QY 298 TTGTGGTTGAATGGTGGCCCTGGA 321
Db 48 LeutrpleuAsnGlyProGly 55
RESULT 15
US-09-901-252-15
; Sequence 15, Application US/09901252
; Patent No. US20020026658A1
; GENERAL INFORMATION:
; APPLICANT: Chapple, Clint
; TITLE OF INVENTION: Genes Encoding Sinapoylglucose:Malate Sinapoyltransferase and Met
; TITLE OF INVENTION: Use
; FILE REFERENCE: NL422-004
; CURRENT APPLICATION NUMBER: US/09/901,252
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/216593
; PRIOR FILING DATE: 2000-07-07

; PRIOR APPLICATION NUMBER: US 60/216593
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Saccharomyces
US-09-901-252-17
Alignment Scores:
Pred. No.: 130 Length: 421
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 9 Gaps: 0
US-09-712-338-1 (1-1668) x US-09-901-252-17 (1-421)
QY 298 TTGTGGTTGAATGGTGGCCCTGGA 321
Db 48 LeutrpleuAsnGlyProGly 55
Search completed: November 21, 2003, 17:58:41
Job time : 59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2003, 20:31:11 ; Search time 3271 Seconds
(without alignments)
12393.722 Million cell updates/sec

Title: US-09-712-338-1

Perfect score: 1668

Sequence: 1 atgctgtgctacgaattctt.....ccagtgtgtgcatgcatag 1668

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vit:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rtd:*

26: em_gss_pbg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	138.4	8.3	588	9	AW672518	AW672518 LG1_360.F
2	116	7.0	721	14	CA747625	CA747625 GAP13D07
3	92.6	5.6	498	12	BM868983	BM868983 mgns002xm
4	86.6	5.2	963	29	CNS0625E	AL421896 T3 end of

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	86.6	5.2	1049	29	CNS06XWN	AL420285 T3 end of
6	83.6	5.0	696	13	BUE34841	BUE34841 035D05 In
7	80.6	4.8	485	12	BI200653	BI200653 oia06fs.i
8	77.6	4.7	445	12	BI187544	BI187544 a4e08fs.i
9	77.6	4.7	475	9	AW677015	AW677015 DGL_3_C02
10	77	4.6	1054	29	CNS06NN7	AL406985 T7 end of
11	74	4.4	729	13	BQ110016	BQ110016 VD0102C06
12	72.8	4.4	1002	29	CNS06ZOK	AL422586 T3 end of
13	72.4	4.3	662	14	CD044859	CD044859 pshb015xg
14	72.2	4.3	682	14	CD054523	CD054523 HO01B22r
15	70.2	4.2	907	14	CD456151	CD456151 F903.10g0
16	69.2	4.1	675	14	CB937789	CB937789 IPCGJX13
17	68.2	4.1	1109	14	CD508724	CD508724 CDA92-D11
18	68	4.1	580	13	BI197413	BI197413 CD197413
19	67.6	4.1	700	12	BI749802	BI749802 F902 05d0
20	67.4	4.0	685	13	BI194502	BI194502 BW194502
21	66.2	4.0	457	14	CD039988	CD039988 pshb031xm
22	66.2	4.0	544	14	CD042968	CD042968 pshb043xf
23	66.2	4.0	712	14	CD045123	CD045123 pshb017xn
24	66	4.0	648	10	BE777028	BE777028 MY-23-G-0
25	66	4.0	653	14	CD258740	CD258740 D1MY023ag
26	65.8	3.9	482	9	AV985320	AV985320 AV985320
27	65.8	3.9	550	9	AV957236	AV957236 AV957236
28	65.8	3.9	566	13	BW276795	BW276795 BW276795
29	65.8	3.9	686	13	BW260398	BW260398 BW260398
30	65.4	3.9	287	14	W06667	W06667 T2241 MVAT4
31	64.8	3.9	590	13	BI198196	BI198196 BW198196
32	64.2	3.8	551	9	AV976226	AV976226 AV976226
33	64.2	3.8	555	9	AV974977	AV974977 AV974977
34	64.2	3.8	560	12	BP007496	BP007496 BP007496
35	64.2	3.8	585	9	AV956551	AV956551 AV956551
36	64.2	3.8	638	13	BW282449	BW282449 BW282449
37	64.2	3.8	652	9	AV991679	AV991679 AV991679
38	64.2	3.8	656	13	BW276796	BW276796 BW276796
39	64.2	3.8	656	13	BW292110	BW292110 BW292110
40	64.2	3.8	675	13	BW274166	BW274166 BW274166
41	64.2	3.8	675	13	BW302983	BW302983 BW302983
42	64.2	3.8	680	13	BW208298	BW208298 BW208298
43	64.2	3.8	684	13	BW248944	BW248944 BW248944
44	64.2	3.8	685	13	BW251033	BW251033 BW251033
45	64.2	3.8	686	13	BW252952	BW252952 BW252952

ALIGNMENTS

RESULT 1	AW672518	588 bp	mRNA	linear	EST 19-JUL-2000
LOCUS	LG1_360_F03.b1_A002	Light Grown 1 (LGI)	Sorghum bicolor	cdna, mRNA	
DEFINITION	sequence.				
ACCESSION	AW672518				
VERSION	AW672518.1	GI:7536439			
KEYWORDS	EST.				
SOURCE	Sorghum bicolor (sorghum)				
ORGANISM	Sorghum bicolor				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.				
AUTHORS	1 (bases 1 to 588)				
TITLE	Cordonnier-Pratt M.-M., Gingle, A., Marsala, C. and Pratt, L.H.				
JOURNAL	An EST database from Sorghum: light-grown seedlings				
COMMENT	Unpublished Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Sequences have been trimmed to exclude polyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.				

398 CAGACGCGCTACCGCATCTTCAACCGCGCCTGTTCGACAGGACGTGCGCGCGCGCTG 457

1558 AAGAAGATCTGGCCACGCTACAA 1580

458 GTGCACACGGCGACGACACCCGGA 480

RESULT 4
 CINS0625E
 LOCUS
 DEFINITION
 T3 end of clone AY0AA015G04 of library AY0AA from strain CBS 6340 of Kluyveromyces thermotolerans, genomic survey sequence.

ACCESSION
 AL421896
 VERSION
 AL421896.1 GI:12205091
 KEYWORDS
 GSS.

SOURCE
 ORGANISM
 Kluyveromyces thermotolerans
 Kluyveromyces thermotolerans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

REFERENCE
 1 (bases 1 to 963)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Bröttier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekai,F., Toffano-Nicche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)

TITLE
 MEDLINE
 PUBLISHED
 20584711
 11152876
 2 (bases 1 to 963)
 Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. and Dujon,B.
 Genomic exploration of the hemiascomycetous yeasts: 10.
 Kluyveromyces thermotolerans
 FEBS Lett. 487 (1), 61-65 (2000)

JOURNAL
 MEDLINE
 PUBLISHED
 20584720
 11152885
 3 (bases 1 to 963)
 Genoscope.
 Direct Submission

REFERENCE
 1
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Kluyveromyces angustis, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES
 source
 1..963
 /organism="Kluyveromyces thermotolerans"
 /mol_type="genomic DNA"
 /strain="CBS 6340"
 /db_xref="taxon:4916"
 /clone="AY0AA015G04"
 /clone_lib="AY0AA"
 /note="end : T3"
 <270..>962
 /note="similar to Saccharomyces cerevisiae ORF YER139W [strong similarity to carboxypeptidase]"
 /evidence=not_experimental

misc_feature
 242 a 231 c 217 g 272 t 1 others

BASE COUNT
 242 a 231 c 217 g 272 t 1 others

ORIGIN
 Query Match 5.2%; Score 86.6; DB 29; Length 963;
 Best Local Similarity: 55.15%; Score 55.15; DB 29; Length 963;

keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source

1.1049

/organism="Kluyveromyces thermotolerans"

/mol_type="genomic DNA"

/strain="CBS 6340"

/db_xref="taxon:4916"

/clone="AY0AA004D06"

/clone_lib="AY0AA"

/note="end : T3"

<82..>1020

/note="similar to Saccharomyces cerevisiae ORF YBR139w [strong similarity to carboxypeptidase]"

/evidence=not experimental

misc_feature

<172..>1017

/note="similar to Saccharomyces cerevisiae ORF YMR297w [PRC1 ; carboxypeptidase Y, serine-type protease]"

/evidence=not experimental

BASE COUNT 262 a 253 c 252 g 280 t 2 others

ORIGIN

Query Match 5.2%; Score 86.6; DB 29; Length 1049;

Best Local Similarity 55.5%; Pred. No. 9.9e-14;

Matches 167; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 198 TGTCAATCTTACTCTGATATGCGACACCTCTCCGAGTCCCATACCTTCTCTGGTT 257

Db 434 TGTCAAGAGTACTCGGGTATCTGAGTACGAGATCCAAACATCTTTTACTGGC 493

QY 258 CTTTCAAGCCAGACATACCCAGAACTGCACTATACATGTTGGTTGAATGGTGGCC 317

Db 494 ATTGGAATCCAGAAACGACCCACTCAATGACCCGTTATTTCTATGGCTGAACGGTGGTCC 553

QY 318 TGAAGCGATCTTTGATCGGTCTCTTCGAAGTGGCCCTTCCCATGTCATTTCCAC 377

Db 554 AGGCTGCTCTTCTTTCACGGGTTATTTTTCGAATTTGGCCCTTCTCCGTCGGCCCTGA 613

QY 378 TTTTGTGATCATCACTCACTCTGCTGGAACGAGTCTCCAAATTTACTATTCCTGTC 437

Db 614 GCTAAAGCCCTTCCGATTCGCTCTCTGGAATACATGCTACTGTCATATTTTGA 673

QY 438 CCAGCCATGGGAGTGGCTTTTTCATATGATGATACGTTGATGGGTCCATTAACCTGT 497

Db 674 GCAGCCCTTAGCGTGGTGGTTTCTCTATGCTGACGAGAGTTGCTCCACTAACGCAGC 733

QY 498 A 498

Db 734 A 734

RESULT 5

CNS06XWN

LOCUS

DEFINITION

T3 end of clone AY0AA004D06 of library AY0AA from strain CBS 6340 of Kluyveromyces thermotolerans, genomic survey sequence.

ACCESSION

AL420285

VERSION

AL420285.1 GI:12203469

KEYWORDS

GSS.

ORGANISM

Kluyveromyces thermotolerans

Kluyveromyces thermotolerans

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

REFERENCE

AUTHORS

Souci t,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brot tier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuv glise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tek ia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weiss nbach,J.

TITLE

Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL

MEDLINE

FEBS Lett. 487 (1), 3-12 (2000)

REFERENCE

AUTHORS

2 (bases 1 to 1049)

Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. and Dujon,B.

TITLE

Genomic exploration of the hemiascomycetous yeasts: 10.

JOURNAL

MEDLINE

FEBS Lett. 487 (1), 61-65 (2000)

REFERENCE

AUTHORS

3 (bases 1 to 1049)

Direct Submission

TITLE

Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cr mieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

JOURNAL

MEDLINE

FEBS Lett. 487 (1), 61-65 (2000)

REFERENCE

AUTHORS

COMMENT

This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See

RESULTS

BU634841 696 bp mRNA linear EST 23-SEP-2002

035D05 Infected Arabidopsis Leaf Arabidopsis thaliana cDNA, mRNA

sequence.

BU634841

VERSION

BU634841.1 GI:23302096

KEYWORDS

EST.

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE

AUTHORS

1 (bases 1 to 696)

Lundsgaard,M., Emmersen,J., Nielsen,K.L., Wilson,I., Somerville,S. and Wellinder,K.G.

TITLE

EST sequencing of Erysiphe cichoracearum infected Arabidopsis plants

JOURNAL

COMMENT

Unpublished

Contact: Karen G. Wellinder

Institut for bioteknologi

Aalborg Universitet

Sohngaardsholmsvej 49, 9000 Aalborg, Denmark
 Tel: +45 96358467
 Fax: +45 98141808
 Email: kgw@bio.auc.dk.

FEATURES
 source Location/Qualifiers

1..696
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /dev_stage="Plant 3 weeks old, three days post infection"
 /clone_lib="Infected Arabidopsis Leaf"
 /note="Organ: Leaf; Vector: pBluescript; Mixed cDNA library of Arabidopsis and E. chichoracearum infected leaf from three weeks old Arabidopsis plants. Plants were harvested 3 days after infection and mRNA oligo dT selected."

BASE COUNT 207 a 129 c 160 g 200 t
 ORIGIN

Query Match 5.0%; Score 83.6; DB 13; Length 696;
 Best Local Similarity 58.5%; Pred. No. 5.8e-13;
 Matches 165; Conservative 0; Mismatches 114; Indels 3; Gaps 1;

QY 175 GGCGTCTGCAGACTACCCGGGTGTCAAATCTCTCTGATATGTCGACACCTCTCCC 234
 Db 146 GGTATCTGTGAGACGACAGAGGTGTTAACCAATATTCTGGATATCTTCGGATAGTAT 205
 QY 235 GAGTCCCATACCTTCTCTGTTCTTCGAAGCCAGACATACCCAGAACTGCACCTATC 294
 Db 206 GATGAGAACATGTGTTCTCTGTTCTTCGAGGCTAGAAATAATCCAATACGGCACCAGTA 265
 QY 295 ACATTGTGTTGAATGTGGCCCTCGAAGCGATTCTTTGATCGTCTCTCTCGAGAGTTG 354
 Db 266 GTGGTGTGGTCAATGGGGCCCGCGTCTCATCATGATGCTTAATTCGAAGAAAT 325
 QY 355 GGCCCTTGCA--TGTCAATCGACTTTTGATGACTATACATCAACCTCTACTCGTGAAC 411
 Db 326 GGTCTTGTTCATTTGAAGATGATCCATGAACCATCCCTCAACCGTACAGCTGAAT 385
 QY 412 GAGGTCTCAATTTACTATTCTGTCCAGCCATTGGGAGTC 453
 Db 386 GAATTGGCAACATGCTCTACATGACCAAGCTTTCAATC 427

RESULT 7
 BI200653
 LOCUS 485 bp mRNA linear EST 10-JUL-2001
 DEFINITION library Fusarium sporotrichioides Tri 10 overexpressed cDNA sequence.

ACCESSION BI200653
 VERSION BI200653.1 GI:14666625
 KEYWORDS EST.

SOURCE Fusarium sporotrichioides
 ORGANISM Fusarium sporotrichioides

REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocremetidae; Hypocreales; mitosporic Hypocreales; Fusarium. 1 (bases 1 to 485)
 AUTHORS Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand ,M. and Roe,B.

TITLE Analysis of a Fusarium sporotrichioides EST database
 JOURNAL Unpublished
 COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu

Contact Dr. Marian Beremand regarding clone availability Included
 is the best homolog from a blastx search of Genbank nr 04-09-01
 466 2e-46 gi|461830|sp|P34946| CARBOXYPEPTIDASE Sipir||S38953

carboxypept
 Seq primer: T3
 High quality sequence stop: 467.
 Location/Qualifiers

1..485
 /organism="Fusarium sporotrichioides"
 /mol_type="mRNA"
 /strain="Tri 10"
 /db_xref="taxon:5514"
 /clone_lib="olao6fs"
 /cdna_library="Fusarium sporotrichioides Tri 10 overexpressed cDNA library"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript ; 3' end of cDNA cloned into XhoI site of pBluescript"
 BASE COUNT 122 a 128 c 120 g 115 t
 ORIGIN

Query Match 4.8%; Score 80.6; DB 12; Length 485;
 Best Local Similarity 57.5%; Pred. No. 3.5e-12;
 Matches 165; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

QY 175 GGCGTCTGCAGACTACCCGGGTGTCAAATCTCTCTGATATGTCGACACCTCTCCC 234
 Db 177 GGAATATCGAGACTACCGCAGGTGTCGCCAGCACTCTGTTATTTCTCTGCGGAC 236
 QY 235 GAGTCCCATACCTTCTCTGTTCTTTTGAAGCCAGACATACCCAGAACTGCACCTATC 294
 Db 237 AACATGAACATGTGTTCTGTTCTTCGAGTCCCGAAGAACGCCAAACAGCTCCACTG 296
 QY 295 ACATTGTGTTGAATGTGGCCCTCGAAGCGATTCTTTGATCGTCTCTCGAAGATTG 354
 Db 297 GCCTTTGGTCAATGTGGCCAGGCTGTAGTTCATGATCGTCTGTTCCAGGAAAT 356
 QY 355 GGCCCTTGCCATGCTCAATTCG---ACTTTGATGACTACATCAACCTCTACTCGTGAAC 411
 Db 357 GGCCCTTGCCATGCTCAACAGGAGGCTCAAGCCAACTTTAAACCTTACTCATGGAAC 416
 QY 412 GAGGTCTCAATTTACTATTCTGTCCAGCCATTGGGAGTCGGCTT 458
 Db 417 ACATTGCCAACATGTTATATGTTGATCAACCTATTGGCACTGGCTT 463

RESULT 8
 BI187544
 LOCUS 445 bp mRNA linear EST 10-JUL-2001

DEFINITION library Fusarium sporotrichioides Tri 10 overexpressed cDNA sequence.

ACCESSION BI187544
 VERSION BI187544.1 GI:14661223
 KEYWORDS EST.

SOURCE Fusarium sporotrichioides
 ORGANISM Fusarium sporotrichioides

REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocremetidae; Hypocreales; mitosporic Hypocreales; Fusarium. 1 (bases 1 to 445)
 AUTHORS Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand ,M. and Roe,B.

TITLE Analysis of a Fusarium sporotrichioides EST database
 JOURNAL Unpublished
 COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu

Contact Dr. Marian Beremand regarding clone availability Included
 is the best homolog from a blastx search of Genbank nr 04-09-01
 386 5e-37 gi|461830|sp|P34946| CARBOXYPEPTIDASE Sipir||S38953
 carboxypept
 Seq primer: T3

```

FEATURES
  source
    High quality sequence stop: 413.
    Location/Qualifiers
      1..445
        /organism="Fusarium sporotrichioides"
        /mol_type="mrna"
        /strain="Tri 10"
        /db_xref="taxon:5514"
        /clone="a4e08fs"
        /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
        cDNA library"
        /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
        XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
        ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT
  113 a 117 c 109 g 106 t
ORIGIN
  Query Match
  Best Local Similarity 4.7%; Score 77.6; DB 12; Length 445;
  Matches 153; Conservative 0; Mismatches 104; Indels 3; Gaps 1;
  QY 175 GCGCTCTGCGAGTACCCGGGTGTCAAATCCTACTCTGGATATGTCGACACCTCTCC 234
  Db 186 GGAATATGCGAGTACCGCAGGTGTCGCCAGCACTCTGTTATTCTCTGCGGAC 245
  QY 235 GAGTCCCACTCTCTCTGTTCTGAGGCGAGACATAACCCAGAACTGCACTATC 294
  Db 246 AACATGAACATCTGTTCTGTTCTCGAGTCCGCAAGACGCCAAACAGCTCCACTG 305
  QY 295 ACATTGTGTTGAATGGCCCTCGAAGGATCTTTGATCGGTCTCTTCAAGAGTTG 354
  Db 306 GCGCTTTGCTCAATGGTGGCCAGGCTGTAGTTCATGATCGGTCTGTTCAGGAAAT 365
  QY 355 GGCCTTGCACTCAATTCG---ACTTTGATGACTACATCAACCTCACTGCTGAAC 411
  Db 366 GGCCTTGCACTTCAACAGGAGGTCAAGCCAACTTAAACCTTACTCATGGAAC 425
  QY 412 GAGGTCTCCAATTACTATT 431
  Db 426 ACATTCGCCAACATGTTATT 445
RESULT 9
AW677015 475 bp mrna linear EST 19-JUL-2000
LOCUS DGI_3 C02.bi_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mrna
DEFINITION
ACCESSION AW677015.1 GI:7550690
VERSION
KEYWORDS
SOURCE
ORGANISM Sorghum bicolor (sorghum)
REFERENCE
  1 (bases 1 to 475)
  Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
  L.H.
  An EST database from Sorghum: dark-grown seedlings
  Unpublished
  Contact: Cordonnier-Pratt MW
  Laboratory for Genomics and Bioinformatics
  The University of Georgia, Department of Plant Biology
  Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
  Tel: 706 542 1860
  Fax: 706 583 0210
  Email: mmpratt@uga.edu
  Sequences have been trimmed to exclude PolyA, vector and regions
  below Phred quality 16. The threshold for highest quality sequence
  is 20.
  Seg primer: JEN REV
  High quality sequence stop: 370
  POLYA-No. Location/Qualifiers
FEATURES
  source
    High quality sequence stop: 413.
    Location/Qualifiers
      1..445
        /organism="Fusarium sporotrichioides"
        /mol_type="mrna"
        /strain="Tri 10"
        /db_xref="taxon:5514"
        /clone="a4e08fs"
        /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
        cDNA library"
        /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
        XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
        ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT
  113 a 117 c 109 g 106 t
ORIGIN
  Query Match
  Best Local Similarity 4.7%; Score 77.6; DB 12; Length 445;
  Matches 153; Conservative 0; Mismatches 104; Indels 3; Gaps 1;
  QY 175 GCGCTCTGCGAGTACCCGGGTGTCAAATCCTACTCTGGATATGTCGACACCTCTCC 234
  Db 186 GGAATATGCGAGTACCGCAGGTGTCGCCAGCACTCTGTTATTCTCTGCGGAC 245
  QY 235 GAGTCCCACTCTCTCTGTTCTGAGGCGAGACATAACCCAGAACTGCACTATC 294
  Db 246 AACATGAACATCTGTTCTGTTCTCGAGTCCGCAAGACGCCAAACAGCTCCACTG 305
  QY 295 ACATTGTGTTGAATGGCCCTCGAAGGATCTTTGATCGGTCTCTTCAAGAGTTG 354
  Db 306 GCGCTTTGCTCAATGGTGGCCAGGCTGTAGTTCATGATCGGTCTGTTCAGGAAAT 365
  QY 355 GGCCTTGCACTCAATTCG---ACTTTGATGACTACATCAACCTCACTGCTGAAC 411
  Db 366 GGCCTTGCACTTCAACAGGAGGTCAAGCCAACTTAAACCTTACTCATGGAAC 425
  QY 412 GAGGTCTCCAATTACTATT 431
  Db 426 ACATTCGCCAACATGTTATT 445
RESULT 10
CNS06NN7 1054 bp DNA linear GSS 04-JUL-2001
LOCUS T7 end of clone AU0RA015E08 of library AU0AA from strain CBS 3082
DEFINITION
ACCESSION AL406985
VERSION
KEYWORDS
SOURCE
ORGANISM Saccharomyces kluyveri
REFERENCE
  1 (bases 1 to 1054)
  Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
  Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
  de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
  Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
  Saurin,W., Tekai,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
  Wincker,P. and Weissenbach,J.
  Genomic exploration of the hemiascomycetous yeasts: 1. A set of
  yeast species for molecular evolution studies
  FEBS Lett. 487 (1), 3-12 (2000)
  20584711
  11152876
  2 (bases 1 to 1054)
  Neuvéglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
  Gaillardin,C. and Casaregola,S.
  Genomic exploration of the hemiascomycetous yeasts: 9.
  Saccharomyces kluyveri
  FEBS Lett. 487 (1), 56-60 (2000)
  20584719
  11152884
  3 (bases 1 to 1054)
  Genoscope.
  Direct Submission
  Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
  2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail:
  segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
  This GSS is part of a random genomic sequencing program of thirteen
  yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
  
```

exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source

1..1054
 /location/Qualifiers
 /organism="Saccharomyces kluyveri"
 /mol_type="genomic DNA"
 /strain="CBS 3082"
 /db_xref="taxon:4934"
 /clone="AU0AA015E08"
 /clone_lib="AU0AA"
 /note="End : 17"

misc_feature

<148..>1052
 /note="similar to Saccharomyces cerevisiae ORF YMR297w [PRC1 : carboxypeptidase Y, serine-type protease]
 2 putative frameshift(s)"
 /evidence=not experimental

BASE COUNT 272 a 219 c 229 g 329 t 5 others
 ORIGIN

Query Match 4.6%; Score 77; DB 29; Length 1054;
 Best Local Similarity 53.1%; Pred. No. 6.4e-11;
 Matches 164; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 241 CATACCTCTTCTCGTTCGAGCCAGACATACCCGAAATGCACTATGCAATG 300
 |||||
 Db 286 CATTCCTTCTACTGGTTTTTGAAGTAGGAATGACCAAAAATGACCCCTATCGTCTT 345
 |||||
 QY 301 TGGTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGTCTCTTCGAAGAGTGGGCCCT 360
 |||||
 Db 346 TGGTGAAGCGAGGTGCGAGTTGCTCTCTTTGACTGGTTATTTTCGAATGGGTCT 405
 |||||
 QY 361 TGCATGTCATTCGATTTGATGACTACATCAACCCCTCACTCGTGGACGAGGTCTCC 420
 |||||
 Db 406 TCTTCAATTGGCCAGATTGAAACCAATTACAAACCTTACCTTGGATTCCAATGCC 465
 |||||
 QY 421 AATTACTATTCTGTCGAGCCATTCGGAGTGGCTTTTCATATAGTATGACGTTGAT 480
 |||||
 Db 466 TCGTCACTTTTGGACGAGCCAGTTAACGTGGTTATTCTTACTCTTCTCTGAAGGT 525
 |||||
 QY 481 GGTTCATTAACTGCTGAACTGGGCTGCTGAAATTCGAGCTTCGAGAGTTCAGGC 540
 |||||
 Db 526 GTTTCAGACACCTGCTGCGGACAGGATGTATGCTTTCTTCGAGCTGTTTTCAGA 585
 |||||
 QY 541 CGGTACCCA 549
 |||||
 Db 586 CAGTCCCA 594

RESULT 11

B0110016

LOCUS B0110016 729 bp mRNA linear EST 01-APR-2003
 DEFINITION V0102006 VD01 Verticillium dahliae cDNA, mRNA sequence.
 ACCESSION B0110016
 VERSION B0110016.1 GI:29426386

KEYWORDS

SOURCE

ORGANISM

Verticillium dahliae

Verticillium dahliae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; mitosporic Hypocreales;

Verticillium.

1 (bases 1 to 729)

Neumann,M.J. and Dobinson,K.F.

Sequence tag analysis of gene expression during pathogenic growth

and microsclerotia development in the vascular wilt pathogen

Verticillium dahliae

Fungal Genet. Biol. 38 (1), 54-62 (2003)

2242000

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

12553936

PUBMED

COMMENT

Contact: Dobinson KF
 Agriculture & Agri-Food Canada
 1391 Sandford St., London, Canada, N5V 4T3
 Email: dobinsonk@agr.ca
 Plate: 2 row: 67 column: 6.

FEATURES

source

1..729
 /location/Qualifiers
 /organism="Verticillium dahliae"
 /mol_type="mRNA"
 /strain="Dvd-T5"
 /db_xref="taxon:27337"
 /lab_host="E. coli"
 /clone_lib="VD01"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Developing microsclerotia (DMS) cDNA library; constructed from axenic cultures grown on cellulose membranes overlaid onto a basal medium agar. Cells were harvested at 4 days post-inoculation. Single-pass sequencing was done using the T3 promoter primer: 5' ATTAACCTCACTAAGGGA 3'."

BASE COUNT 152 a 243 c 188 g 144 t 2 others
 ORIGIN

Query Match 4.4%; Score 74; DB 13; Length 729;
 Best Local Similarity 58.7%; Pred. No. 3.9e-10;
 Matches 128; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 247 TTCTTCTGGTTCCTCGAAGCCAGACATAACCCGAAATGCACTATGCAATG 306
 |||||
 Db 171 TTCTACTGGTTCCTCGAGTCCCGCAATGACCCCAAGAACGCCGCTCTCTGCGTC 230
 |||||
 QY 307 AATGGTGGCCCTGGAAGCGATTCTTTGATCGTCTCTTCGAAGAGTGGGCCCTTGCAT 366
 |||||
 Db 231 AACGTGGCCCTGGTGTCTTCCCTGACCGGCTCTTCATGAGCTCGGCCCGCCAGC 290
 |||||
 QY 367 GTCAATTCGATTTGATGACTACATCAACCCCTCACTCGTGGACGAGGTCTCCAAATTA 426
 |||||
 Db 291 ATCAACAAGAAGTCGAGATTGTCAACACGAGTGGTCTTGGACACACACGCTTCGGTC 350
 |||||
 QY 427 CTATTCCTGTCGCCAGCCATTCGGAGTCGGCTTTTCATA 464
 |||||
 Db 351 ATCTTCTTGACCGCCTGTGAACGTTGGCTACTCGTA 388

RESULT 12

CNS06ZOK/c

LOCUS

DEFINITION

T3 end of clone AZ0AA003B09 of library AZ0AA from strain CBS 712 of

Kluyveromyces marxianus, genomic survey sequence.

AL422586

VERSION AL422586.1 GI:12205780

KEYWORDS GSS.

SOURCE Kluyveromyces marxianus

ORGANISM Kluyveromyces marxianus

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

REFERENCE 1 (bases 1 to 1002)

Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,

Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,

de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,

Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,

Saurin,W., Tekai,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,

Wincker,P. and Weissenbach,J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

20584711

PUBMED 11152876

REFERENCE 2 (bases 1 to 1002)

Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.

and Dujon,B.

Genomic exploration of the hemiascomycetous yeasts: 12.

JOURNAL Kluveromyces marxianus var. marxianus
 MEDLINE FBS Lett. 487 (1), 71-75 (2000)
 PUBMED 20584722
 REFERENCE 11152887
 3 (bases 1 to 1002)
 Genoscope.
 Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
 JOURNAL seqrefgenoscope.cns.fr
 COMMENT This GBS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluveromyces thermotolerans, Kluveromyces
 lactis var. lactis, Kluveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.
 FEATURES
 Location/Qualifiers
 1..1002
 /organism="Kluveromyces marxianus"
 /mol_type="genomic DNA"
 /strain="CBS 712"
 /variety="marxianus"
 /db_xref="taxon:4911"
 /clone="AZ0AA003B09"
 /clone_lib="AZ0AA"
 /notes="end : T3"
 complement (<8..>994)
 /note="similar to Saccharomyces cerevisiae ORF YMR297w [
 PRC1 ; carboxypeptidase Y, serine-type protease]"
 /evidence="not experimental"
 complement (<11..>778)
 /note="similar to Saccharomyces cerevisiae ORF YBR139w [
 strong similarity to carboxypeptidase]"
 /evidence="not experimental"
 BASE COUNT 267 a 259 c 240 g 235 t 1 others
 ORIGIN
 Query Match 4.4%; Score 72.8; DB 29; Length 1002;
 Best Local Similarity 53.5%; Pred. No. 1.1e-09;
 Matches 152; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
 QY 241 CATACCTTCTCTGGTTCCTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTG 300
 Db 673 CACTTCTTCTACCTGGTCTTTGAGTCAGAAACACCCCAAGACGACCTGTATCTTG 614
 QY 301 TGGTTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTTCGAAGAGTTGGGCCCT 360
 Db 613 TGGTTGAAGCGTGGCCAGGCTGTCTCTTGAAGTGGTGTCTTCGAATGGGTCCA 554
 QY 361 TGGCATGTAATTCGACTTTGATGACTATCAACCTCAGCTCGTGAAGAGGTCTCC 420
 Db 553 TCTTCCATTCGCGAAGAGGTGAAGCCCAATTACAAACCACTCTTGAAGAGCAAGCT 494
 QY 421 AATTACTATTTCCTGCCAGCCATTCGGAGTCGGCTTTTCATATAGTGTACCGTTGAT 480
 Db 493 TCCGTTATCTTTGGACAGCAGCTCAACCTTGGTTACTTCTTCTCATCAGAAGGT 434
 QY 481 GGGTCCATTAAACCTGTAACCTGGGGTGTGCGAAATTCGAGCTT 524
 Db 433 GTCTCAACACTGTGTGCTGAGTAAGGACGTGTATGCGTCTT 390
 RESULT 13
 CD044859
 LOCUS
 DEFINITION psHB016xG08f 183244 psHB: Infected hypocotyl soybean host. 48 hrs
 post infection Phytophthora sojae cDNA clone SHB016G08 5, mRNA
 sequence.
 ACCESSION CD044859
 CD044859
 EST.
 Phytophthora sojae
 SOURCE Phytophthora sojae
 ORGANISM Eukaryota; stramenopiles; Comycetes; Pythiales; Pythiaceae;
 Phytophthora.
 1 (bases 1 to 662)
 Tyler B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
 USDA-IFAFS: Expression of Phytophthora sojae genes during infection
 and propagation
 Unpublished
 Contact: Tyler B
 Tyler lab
 VBI
 1860 Pratt Dr., Blacksburg, VA 24061, USA
 Tel: 540-231-7318
 Email: bmtyley@vt.edu
 PCR Primers
 FORWARD: BK reverse
 Plate: 016 row: G column: 08
 Seq primer: BK reverse stop: 662.
 High quality sequence stop: 662.
 Location/Qualifiers
 1..662
 /organism="Phytophthora sojae"
 /mol_type="mRNA"
 /db_xref="taxon:67593"
 /clone="SHB016G08"
 /tissue_type="infected host tissue"
 /cell_line="P6497"
 /dev_stage="48 hour post infection"
 /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs
 post infection"
 /note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI;
 USDA-IFAFS: Expression of Phytophthora sojae genes during
 infection and propagation."
 BASE COUNT 130 a 204 c 208 g 120 t
 ORIGIN
 Query Match 4.3%; Score 72.4; DB 14; Length 662;
 Best Local Similarity 53.0%; Pred. No. 1.1e-09;
 Matches 179; Conservative 0; Mismatches 156; Indels 3; Gaps 1;
 QY 142 GTCACCATCGGTACAGGAACCCGGGGCAGAGGGCTCTCGAGACTACCCGGGTGTC 201
 Db 217 GTCACCTAACAGTCTGAGACCCGACAGATTCTCTGCGGATTACGAATCACGAG 276
 QY 202 AAATCCTTACTCTGGATATGTCGACACCTCTCCGAGTCCCAACCTTCTTCTGTTCTTC 261
 Db 277 ACGGGCTACATTAAAT---TACCCACAAGGACGACCACTACTTCTACTGTTCTGTC 333
 QY 262 GAAGCCAGACATAACCCAGAACTGCACCTATCACATTGTTGATGTGGCCCTGGA 321
 Db 334 GAGTCGCGCAGCAGTCGCGAAGGACCCCTTTGGTGTCTGGCTCAGGGCGGCCCGGC 393
 QY 322 AGCGATTTCTTGTGCTCTCTTCGAAGAGTGTGGCCCTTGGCCATGTCATTTGACTTTT 381
 Db 394 TGCTCCAGCAATGAGCGCTACTGCGCGAATGTTCTTGGCAGCGGAGGAGCTG 453
 QY 382 GATGACTATCATCAACCTCTACTCGTGAAGAGGTCTCCAAATTTACTATTCTCTCCAG 441
 Db 454 TCCACCAAGACCAACCCGCTACTCGTGAAGCGGCAATGTTATTGGCTGGACCAA 513
 QY 442 CCATTGGAGTCGGCTTTTCATATAGTATAGTACGGTTGA 479
 Db 514 CCCACAGCGGTGGGCTACTCGTACGGCCCGCCAGGTGGA 551
 RESULT 14
 CD054523
 LOCUS
 DEFINITION CD054523 HO Hordeum vulgare cDNA clone HO01B22 5-PRIME, mRNA
 sequence.
 CD054523
 EST
 mRNA
 linear
 682 bp
 CD054523
 LOCUS
 DEFINITION CD054523 HO Hordeum vulgare cDNA clone HO01B22 5-PRIME, mRNA
 sequence.
 ACCESSION CD054523

ACCESSION CD054523
 VERSION CD054523.1 GI:30595544
 KEYWORDS EST.
 SOURCE Hordeum vulgare
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 682)
 AUTHORS Zierold, U. and Schweizer, P.
 TITLE Barley ESTs from pathogen-attacked leaf epidermis
 JOURNAL Unpublished
 COMMENT Contact: Patrick Schweizer
 Transcriptome Analysis, Cytogenetics Department
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3, D-06466 Gatersleben, Germany
 Tel: 0049 (0)39482-5660
 Fax: 0049 (0)39482-5595
 Email: schweiz@ipk-gatersleben.de
 Insert Length: 682 Std Error: 0.00
 Plate: 1 row: B column: 22
 Seq primer: M13rev.

FEATURES
 source
 1..682
 Location/Qualifiers
 /organism="Hordeum vulgare"
 /mol_type="mRNA"
 /cultivar="Ingrid BC mlo-5"
 /db_xref="GABI:703742"
 /db_xref="taxon:4513"
 /clone="H001822"
 /tissue_type="leaf epidermis, 6 h and 24 h post inoculation with Blumeria graminis"
 /dev_stage="7 d after germination"
 /lab_host="XL10-Gold"
 /clone_lib="HO"
 /note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Approximately 5 % of the clones correspond to cDNA from the fungi B. graminis hordei and tritici, respectively. Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1.2 kb"

BASE COUNT 191 a 173 c 152 g 166 t
 ORIGIN

Query Match 4.3%; Score 72.2; DB 14; Length 682;
 Best Local Similarity 57.2%; Pred. No. 1.3e-09;
 Matches 131; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
 QY 241 CATACCTCTCTCGGTTCTTCGAGCCAGACATACCCGAACTGCACCTATCATG 300
 Db 438 GATCTCTTTTACGGTTTTTTGAGTCCCGAAACGATCCCGAGAACCCAGTCATCTG 497
 QY 301 TGGTTGAATGGTGGCCCTGGAGCGGATCTTTGATCGGTCTCTTCGAGAGTTGGCCCT 360
 Db 498 TGCTCAACGGCGCCCTGGATGTTCTGTCATTAACCTGGATTAATCTTGGAGTCTGGCCCG 557
 QY 361 TGGCATGTCAATTCGACTTTTGTGACTATCATCAACCTCTCATCTCGTGGAAACGAGGTCCTC 420
 Db 558 GCCTCAATCGATAAAACCTTGAAGGTGGTCAATAACCCCTATTCCTGGAACTGAACGCA 617
 QY 421 AATTACTATTCCTGTCCAGCCATGGGAGTCGGCTTTTCATPATAGT 469
 Db 618 TCAGTTATTTTCTTGACCAACCTGTAAATGTCGGATCTCGTACAGCG 666

RESULT 15
 CD456151/c 907 bp mRNA linear EST 03-JUN-2003
 LOCUS CD456151

DEFINITION

Fg03_10g07_R
 Fg03_AAFRC_ECORC_Fusarium_graminearum_mycelium_trichothecene_product
 ion Gibberella zeae cDNA clone Fg03_10g07, mRNA sequence.

ACCESSION

CD456151

VERSION

CD456151.1 GI:31370891

KEYWORDS

EST.

SOURCE

Gibberella zeae

ORGANISM

Gibberella zeae

REFERENCE

1 (bases 1 to 907)

AUTHORS

Harris, L.J., Allard, S., Rochleau, H., Chapados, J., Couroux, P., Sprott, D., and Finkler, N.A.

TITLE

A cDNA library prepared from Fusarium graminearum grown in liquid culture under trichothecene-producing conditions

JOURNAL

Unpublished

COMMENT

Contact: Harris, Linda J.
 Eastern Cereal and Oilseed Research Centre
 Agriculture and Agri-food Canada
 Bldg. 21, Central Experimental Farm, Ottawa, Ontario, KIA 0C6,
 CANADA

FEATURES

1..907
 Location/Qualifiers
 /organism="Gibberella zeae"
 /mol_type="mRNA"
 /strain="DAOM 180378"
 /db_xref="taxon:5518"
 /clone="Fg03_10g07"
 /tissue_type="Mycelium"
 /dev_stage="Asexual"
 /lab_host="E. coli SURE2"
 /clone_lib="Fg03_AAFRC_ECORC_Fusarium_graminearum_mycelium_trichothecene_production"
 /note="Vector: pGEM-T easy; Site 1: EcoRI; Fusarium graminearum grown in liquid culture to optimize trichothecene production; final deoxynivalenol concentration in culture determined by CD-ELISA to be 165 micrograms/ml. cDNA was prepared using a modified SSH protocol to promote normalization whereby cDNA was not digested prior to hybridization and no subtraction was attempted"

BASE COUNT

243 a 192 c 225 g 242 t

ORIGIN

Query Match 4.2%; Score 70.2; DB 14; Length 907;
 Best Local Similarity 45.9%; Pred. No. 5.8e-09;
 Matches 379; Conservative 2; Mismatches 410; Indels 34; Gaps 3;

QY

724 TTTCAGCAGCAATGAGAGATTGCCAACGGTAGTGTAAATGGTCTTCAGCTTAATTC 783

Db

906 TTTCAGCAGCAATGAGAGATTGCCAACGGTAGTGTAAATGGTCTTCAGCTTAATTC 847

QY

784 AACTCTCTGGGAATTATTAACCGCATCATCGAGCGGATCCAGGCCCTTACTACCT 843

Db

846 GTAGCTCTTGGCATCAACAATGGATGATGATCTCTGCCAACAGTACAGATTACATC 787

QY

844 GAATTCGCTGTCAACATACCTACGGTATCAAGGTGTCAACGAGACCGTCTACAACTAC 903

Db

786 GAATACGCGAGCAATTAACACCTTACAAAAGTGTATGATCACTCAACTCAGTACAGCAAGTAC 727

QY

904 ATGAAGTTTGGCAACCAATGCCAATGTTGCCAGGATTTGATTTCCACCTGGCAACAG 963

Db

726 CTAACACCTCAATCAAAAG-----TGTGTGCTCTTTCGCS 689

QY

964 ACAAAACCGCAGCGCATTTAGCTGACTACGCCCTCTCGCGGAGCCACCAACATGTCGAG 1023

Db

688 AAGTCCCGAGGCTCACAGGCAACCGCTGTTGTGGCATGCCAGCATGCTTGCACT 629

QY

1024 GCAATGTTGGGGGCCATATACGCCCTTGTGCTGTGTGTATGATATTCGGCAT 1083

Db 628 CAGGCTATTGAACGGCCCTTGAGAAATACAGAGACTTTGACGTTTATGATATCGGGCC 569
QY 1084 CCATATGATGACCCGACCTCCGCAAGTTATTACAAATAATTTCTGGCAAGGACTCTGTG 1143
Db 568 CCGAGTAACGACCCCTTTCTCCATCGACTTACTCGACTTATCTTCAGTCAAGTGTG 509
QY 1144 ATGGACGCTATCGGCGTCAACATCAACTACACCCAGTCCCAATPAATGACGTCTACTACGCT 1203
Db 508 ATGAAGCTATTGGCGCGGAGTCAACATACGGGGAATGTCCCGAGGGGCTTATGACAAG 449
QY 1204 TTCCAGCAAAAGGCGACTTTGTCTGGCCCAACTTCTGAGAGACCTCGAGGATCCTT 1263
Db 448 TTTATCAACAGTGGCGATCGGGGACGATCAITCTTTGTCCACATTTGCCAAG---TCATT 392
QY 1264 GCTCTCCCGTGGTGTCTCTCCCTCATCTATGCGAGCGGATTACATCTGCAACTGGTTC 1323
Db 391 GACTCCAGATCCAAGTATTGATCTGGGCTGGTGATGCCGACTGGATCTGCACTGGAT- 333
QY 1324 GCGGCTCAGGCGGTTTCCCTCGTCCGAACACTACTCCCAAGCCGCCAGTTCGGAAGCGCA 1383
Db 332 -----GGGTAATTACCGGCTCTCAACTCTATTGCTCCCGAGTCGTTCTCAGCT 281
QY 1384 GGGTACACGCCCTGAAAGTCAACGCGTCGAGTATGGGAACTCGCGAGTATGTAAT 1443
Db 280 CCTCTTCAATCCCTTTACTGTGATGGAACCTAAGTACGGAGATTTAAGACATCTGGAAT 221
QY 1444 TTCTCCTTCACTCGGCTCTATGAGGCGAGGCGCATGAAGTCCCAATACACAGCCCATCGCC 1503
Db 220 TTGAGCTGGCTGGAGTTTATGGTGTGTCATGAGTCCCGCTTATCAGCTCAGGCT 161
QY 1504 TCCTGCAATTTTAAACCGGACTATCTTGGTTGGGATATGCA 1548
Db 160 GCATTGGCTGCTTTTCTGGCCACACTTTCAGAGAGCTATTTC 116

Search completed: November 21, 2003, 14:18:13
Job time : 3282 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 21, 2003, 17:25:24 ; Search time 86.5 Seconds

(without alignments)
9552.179 Million cell updates/sec

Title: US-09-712-338-1

Perfect score: 3027

Sequence: 1 atcggtggtacgaattctt.....ccagtggtgtaggcacatg 1668

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q-/cgn2.1/USPTO.spool.p/US09712338/runat_17112003.170149.10519/app_query.fasta_1.1963
-DB=SPTRMBL23 -QFMT=FASTAN -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CNC 1.1.127 @runat_17112003.170149.10519 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL 23:*
1: sp.archaea.*
2: sp.bacteria.*
3: sp.fungi.*
4: sp.human.*
5: sp.invertebrate.*
6: sp.mammal.*
7: sp.mhc.*
8: sp.organella.*
9: sp.phage.*
10: sp.plant.*
11: sp.rodent.*
12: sp.virus.*
13: sp.vertebrate.*
14: sp.unclassified.*
15: sp.virus.*
16: sp.bacteriaph.*
17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2976	98.3	555	3	Q96VZ9 aspergillus

ID	Q96VZ9	PRELIMINARY;	PRT;	555 AA.
AC	Q96VZ9			
DT	01-DEC-2001 (Tremblrel. 19, Created)			
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DT	01-VAR-2003 (Tremblrel. 23, Last annotation update)			
DE	Carboxypeptidase SI precursor (EC 3.4.16.6).			
GN	CPI.			
OS	Aspergillus oryzae.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;			
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.			
OX	NCBI_TaxID=5062;			
RN	[1]_TaxID=5062;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=TK3;			
RA	van den Broek P.;			
RL	Submitted (JUN-2001) to the EMBL/GenBank/DDJ databases.			
DR	EMBL; AF394242; AAK77166.1; --			
DR	InterPro; IPR001563; Serine carboxypeptidase.			
DR	Pfam; PF00450; serine carboxypeptidase; 1.			
DR	PRINTS; PR00724; CRBOXPTASEC.			
DR	ProDom; PD001189; Serine carboxypeptidase; 1.			
DR	PROSITE; PS00131; CARBOXYPEPTIDASE; 1.			
KW	Carboxypeptidase; Hydrolase; Signal.			

ALIGNMENTS

RESULT 1

Q96VZ9 ID Q96VZ9 PRELIMINARY; PRT; 555 AA.

AC Q96VZ9
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-VAR-2003 (Tremblrel. 23, Last annotation update)
DE Carboxypeptidase SI precursor (EC 3.4.16.6).
GN CPI.

OS Aspergillus oryzae.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OX NCBI_TaxID=5062;

RN [1]_TaxID=5062;

RP SEQUENCE FROM N.A.

RC STRAIN=TK3;

RA van den Broek P.;

RL Submitted (JUN-2001) to the EMBL/GenBank/DDJ databases.

DR EMBL; AF394242; AAK77166.1; --

DR InterPro; IPR001563; Serine carboxypeptidase.

DR Pfam; PF00450; serine carboxypeptidase; 1.

DR PRINTS; PR00724; CRBOXPTASEC.

DR ProDom; PD001189; Serine carboxypeptidase; 1.

DR PROSITE; PS00131; CARBOXYPEPTIDASE; 1.

KW Carboxypeptidase; Hydrolase; Signal.

FT SIGNAL 1 18 POTENTIAL.
SQ SEQUENCE 555 AA; 61168 MW; 456B63B0CEB55222 CRC64;

Alignment Scores:
Pred. No.: 6,18e-204 Length: 555
Score: 2976.00 Matches: 554
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.82% Mismatches: 0
Query Match: 98.32% Indels: 0
DB: 3 Gaps: 0

US-09-712-338-1 (1-1668) x Q96VZ9 (1-555)

QY	1	ATGCGTGGCTACGAATTTCTCTCAGTGTCTACCTTGGTTCAGCCAGTGGGCCCTTCCA	60
DB	1	MetArgGlyTyrGluPheLeuSerValLeuProLeuValAlaAlaSerTirAlaLeuPro	20
QY	61	GGAGTACACCGCGTCCGTGGTAGACAGACAGTACCCAGAACCCACCGGGTCAAG	120
DB	21	GlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGlyValLys	40
QY	121	ACTCTTACACCGCAACATCTCACCATCCGGTACAGGAACCCGGGCGAGGGCGTC	180
DB	41	ThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyVal	60
QY	181	TGCGAGACTACCCCGGGTGTCAATCTCTACTCTGGATATGTGACACCTCTCCCGAGTCC	240
DB	61	CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSer	80
QY	241	CATACCTTCTCTGGTCTTCGAGCGACAGATACCCAGAACTGCACCTATCACATTG	300
DB	81	HisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu	100
QY	301	TGTTGAATGGTGGCCCTCGAAGCGATTCTTTCATCGCTCTCTTCGACAGTGGGCCCT	360
DB	101	TrpLeuAsnGlyGlyProGlySerSerLeuIleGlyLeuPheGluLeuGlyPro	120
QY	361	TGCGATGTCAATTCGACTTTTGATGACTATACATCAACCTCCTCTCTGCGAACGAGTCTCC	420
DB	121	CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer	140
QY	421	AATTTACTATCTGTCCAGCCATTCGGAGTGGGCTTTTCATATAGTATAGTATCGTGTAT	480
DB	141	AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrAsnAspThrValAsp	160
QY	481	GGTCCATTAACCTGTAACTGGGTCTGCGAAATTCGAGCTTTCGAGGAGTTCAGGGC	540
DB	161	GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly	180
QY	541	CGGTACCCCAACCATTTGATGCCACTCTGATCGATACCAATCTTTCGCGAGAGCGGCT	600
DB	181	ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAla	200
QY	601	TGGAGATCTCTGCAAGATTCCTTAGTGACTACCTAGCTTGGACTTAGGGTGCAGTCT	660
DB	201	TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer	220
QY	661	AAGGACTTTCAGTCTATCGACGAGAGTATGAGGGGCACTATGCTCTGCAATCTTCAAT	720
DB	221	LysAspPheSerLeuThrThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn	240
QY	721	CATTTTACGACGAGATAGAGAAATTCGAACGGTAGTGTATATGTGTTCAGCTTAAT	780
DB	241	HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn	260
QY	781	TTCAACTCTCTGGGAATATTAACGGCATCATCGAGGGCGATCCAGGCCCTTACTAC	840
DB	261	PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr	280
QY	841	CCTGAATTCGTGTGAACATACCTACGGTATCAAGGCTGTCAACGAGACCGCTCAAC	900
DB	281	ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn	300

QY	901	TACATGAAGTTTGCACCAACCAATGCCAAATGGTTCGCCAGGATTGATTTCACCTGCAAA	960
DB	301	TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys	320
QY	961	CAGACAAACCGCACCGCATTTAGCTGACTACACCCCTCTCGCGCCAGCCACCAACATGTGC	1020
DB	321	GlnThrAsnAsnThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys	340
QY	1021	AGGACAAATGTGAGGGGCCATACACCCCTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT	1080
DB	341	ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg	360
QY	1081	CATCCATATGATGACCCGACTCCGCAAGTATTACAAATTTCTGCAAGGACTCT	1140
DB	361	HisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSer	380
QY	1141	GTCATGGACGCTATCGCGCTCAACATCAACTACACCCAGTCCAATTAATGACTCTACTAC	1200
DB	381	ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyr	400
QY	1201	GCTTTCCGCAACAGGCGACTTTGTCTGGCCCACTTCATCGAAGACCTCGAGGATC	1260
DB	401	AlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle	420
QY	1261	CTTGCTCTCCCGTGGTCTCTCCCTCATCTATGCGAGCGCGATTACATCTGCAACTGG	1320
DB	421	LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysIleThr	440
QY	1321	TTCCGCGGTTCAGCGCTTTCCTCTCGCAACTACTCCCAAGCCGCCAGTTCGGAAGC	1380
DB	441	PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer	460
QY	1381	GCAGGTACACCCCTCGAAGTCAACGCGTCTGAGTATGGGAACTCGCAGATATGCT	1440
DB	461	AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly	480
QY	1441	AAATTTCTCTCTCACTCGCTCTATGAGGCGAGCGCATGAAGTCCATCTACACAGCCCATC	1500
DB	481	AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle	500
QY	1501	GCCTCCTCTCAATTTTAAACCGGACTATCTTGGTGGGATATCCAGAGGGCCAGAG	1560
DB	501	AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys	520
QY	1561	AGATCTGCGCCAGCTACAGACCATGAAACGGCTACAGCTACGCTACGCTACGCTACGCT	1620
DB	521	LysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer	540
QY	1621	GTCCCGCTGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCT	1665
DB	541	ValProLeuProThrAlaThrSerMetSerSerValGlyMetAla	555

RESULT 2

ID	Q12569	PRELIMINARY;	PRT;	460 AA.
AC	Q12569;			
DT	01-NOV-1996	(TrEMBLrel. 01, Created)		
DT	01-NOV-1996	(TrEMBLrel. 01, last sequence update)		
DT	01-JUN-2002	(TrEMBLrel. 21, last annotation update)		
DE	Prepro-carboxypeptidase Z (EC 3.4.16.5).			
GN	SPCZ.			
OS	Absidia zychoae.			
OC	Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;			
OC	Absidia.			
OX	NCBI_TaxID=36079;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NEIC 1193;			
EX	MEDLINE=95308533; PubMed=7788719;			
RA	Lee B.R., Takeuchi M., Kobayashi Y.;			
RT	"Molecular cloning and sequence analysis of the scpZ gene encoding the			
RT	serine carboxypeptidase of Absidia zychoae."			
RL	Curr. Genet. 27:159-165(1995).			

QY	814	GACGAGCGATCCAGGCCCTTACCTACCTCGAATTTCGGTGTGAACAATACCTACCGGTATC	873
Db	229	AspAlaValIleGlnTyrLysSerTyrProLysMetThrCysHisSerThrTyrProAla	248
QY	874	AAGCGTGTCAACGAGACCGTCTCAACACTACATGATGAAGTTTGCCAACCAAAATGCCAAATGGT	933
Db	249	ValLeuSerGluGlu-----GluCysAspLysMetGlnGlnIleTyrGluAsnAsp	265
QY	934	TGCCAGGATTTGATTTCCACCTCGCAACAGACAAACCGCACCGCATAGCTGACTACGCC	993
Db	266	CysLysProAlaAlaGluGlnCysAlaGluSerAsp-----GluAspSer	280
QY	994	CTCTGCGCCGAGGCCACCAACATGTGCGAGGACAATGTTGAGGGCCCATACTACGCCGCTTT	1053
Db	281	AspCysValAsnAlaAsnGlnGlnCys---GlyGlnIleGluGly---IleTyrAlaGln	298
QY	1054	GCTGCTCGTGTGTATGATATTCGCATCCATCATGATGACCCGACTCGCCAAAGTTAT	1113
Db	299	SerGlyTyrSerPheTyrAspIleArgGlnGlnGlyAspAsp---ThrProHisProPhe	317
QY	1114	TACAACAAATTTCTGGCAAGAGACTCTGTCATGACGACGCTACGCGGTCAACATCAACTAC	1173
Db	318	ValAspGlu--LeuAsnLysAlaSerValIleLysGluValGlyAlaArgGlyHisPhe	336
QY	1174	ACCAGTCCCAATATGACGTCTACTACGCTTTCAGCAACAGCGCACTTTGTCTGGCC	1233
Db	337	SerMetCysSerAspSerValGlyThrAlaPheAlaGlnThrGlyAspGlyAla---Arg	355
QY	1234	AACTTCATCGAAGACCTCGAGGAGACTCTGCTCTCCCGTGTGCTCTCCCTCATCTAT	1293
Db	356	SerTyrIleProAlaValGluLysLeuLeuLysGluGlyIleProValLeuIleTyrVal	375
QY	1294	GGCGACCCGATACACTCTCAACTGTTCCGGCGGTCAAGCCGCTTCCCTCGCTCGGAAC	1353
Db	376	GlyAspAlaAspValIleCysAsnTyrTyrGlyAsnLeuAspValAlaAspSerLeuLys	395
QY	1354	TACTCCCAACCCGCCGCTTCGAGCGCAGGTACACGCCCTCGAAAGTCAACGGCGCTC	1413
Db	396	TyrAspGlySerAspAlaPheSerLysThrLysLeuGluAlaTyrLysAlaAspGlyLys	415
QY	1414	GAGTATGGGAAACTCGGAGTATGTTTCTCGTCACTCGCGCTCTATGAGGAGGC	1473
Db	416	GluValGlyGlnPheArgSerAlaAspLysLeuThrPheValArgValTyrGluAlaGly	435
QY	1474	CATGAAGTCCCATCTACCAAGCCCATCGCTCCCTCGCAATTTGTTAAACCGCATCTCTC	1533
Db	436	HisGluValProMetTyrGlnProGluAlaLeuSerMetPheGlnThrTrpIleSer	455
QY	1534	GGT 1536	
Db	456	Gly 456	
RESULT	3		
Q96VC4		PRELIMINARY;	PRT; 552 AA.
ID	Q96VC4		
AC	Q96VC4;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DE	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Carboxypeptidase.		
GN	CPYA		
OS	Emmericella nidulans (Aspergillus nidulans).		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;		
OC	Eurotiales; Trichocomaceae; Emericella.		
OX	NCBI_TaxID=162425;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=A26;		
RX	MDLLINE=21333188; PubMed=11440134;		
RA	Ohsumi K., Matsuda Y., Nakajima H., Kitamoto K.;		
RT	"Cloning and characterization of the cpYA gene encoding intracellular		
RT	carboxypeptidase from Aspergillus nidulans."		

RL Biosci. Biotechnol. Biochem. 65:1175-1180 (2001).
 DR EMBL; AB051820; BAB56108.1; -.
 DR MEROPS; S10.001; -.
 DR InterPro; IPR001563; Serine carboxypept.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF00450; serine carboxypept; 1.
 DR PRINTS; PR00724; CRBOXYPTASEC.
 DR ProDom; PD001189; Serine carboxypept; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 SQ SEQUENCE 552 AA; 62065 MW; -F67CF39FDFB7D761 CRC64;

Alignment Scores:
 Pred. No.: 1.95e-34 Length: 552
 Score: 601.00 Matches: 154
 Percent Similarity: 47.34% Conservative: 77
 Best Local Similarity: 31.56% Mismatches: 175
 Query Match: 19.85% Indels: 82
 DB: 3 Gaps: 19

US-09-712-338-1 (1-1668) x Q96VC4 (1-552)

QY 133 GCAACAATGTCACCATCGGTACAGGAACCGGGGAGAGGGGCTCTGCAGACTACC 192
 Db 127 AlaTyrAspLeuArgLysLysThrAspProSerSerLeuGlyLeu-----Asp 143
 QY 193 CGGGGTGTCAAATCTCTACTCTCGATATGTCACACCTCTCCC---GAGTCCCATCTTC 249
 Db 144 ProAspValLysGlnTyrThrGlyTyrLeuAspAspAsnGluAsnAspLysHisLeuPhe 163
 QY 250 TTCTGTTCTTGAACCCAGACATAACCCAGAACTGCACCTATCATCTGTTGTTGAAT 309
 Db 164 TyrTrpPheGluSerArgAsnAspProLysAsnAspProValValLeuTyrLeuAsn 183
 QY 310 GTGGGCTTGAAGCGATCTTTGATCGGTCTCTCGAAGAGTGGGGCTTGGCCATGTC 369
 Db 184 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerLeu 203
 QY 370 AATTCGACTTTTGATGACTACATCAACCTCACTCTGTGAACAGAGTCTCAATTTACTA 429
 Db 204 AspGluAsnLeuLysProValTyrAsnProTyrAlaTrpAsnSerAsnAlaSerValile 223
 QY 430 TTCTGTCCCGCATTTGGAGTCGGCTTTTCATATAGTCATACGGTGTGATGGTCCATT 489
 Db 224 PheLeuAspGlnProValAsnValGlyTyrSerTyrSer----- 236
 QY 490 AACCTCTTAACCTGGGTCGTGCGAAATTCGAGCTTTTCAGGAGTTCAGGGCGGTACCCA 549
 Db 236 ----- 236
 QY 550 ACCATTGATCCCACTCTGATCGATACACCAATCTTCCGCGAGAGCGCGCTTGGAGATC 609
 Db 237 -----GlySerThrValSerAspThrVal---AlaAlaGlyLysAspValTyrAlaLeu 253
 QY 610 CTGCAAGGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGGTGCAGTCTAAGGACTTC 669
 Db 254 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaGluGlnAspPhe 269
 QY 670 AGTCTATGAGCGGAGCATATGGAGGCACATATGGCTTCGATTCATATCAATTTTAC 729
 Db 270 HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheThrSerGluLeuLeu 289
 QY 730 GAGCAGATGAGAGAAATGCCACGGTAGTGTAAATGGTGTTCAGCTTAATTTCACTCT 789
 Db 290 SerHisGlnLysArg-----AsnIleAsnLeuLysSer 300
 QY 790 CTGGGATATTATACGCATCATCGACGAGCGGATCCAGGCGCTTACTACTACCTCGAATTC 849
 Db 301 ValLeuIleGlyAsnGlyLeuThrAspGlyLeuThrGlnTyrGluTyrTyrArgProMet 320
 QY 850 GCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGCGCTACCACTACATCAAGAG 909
 Db 321 AlaCysGlyGluGly---GlyTyrProAlaValLeuAspGlu-----Ser 334

QY 910 TTTGCCAACCAATGCCAAATGGT-----TGCCAGGATTGTGATTTCCACCTGCAAA 960
 Db 335 SerCysArgSerMetAspAsnAlaLeuGlyArgCysGlnSerMetIleGluSerCysTyr 354
 QY 961 CAGACAAACCGCACCGCATTAGCTACGACCTCTCGCCGGAAGCCACCAACATGTC 1020
 Db 355 AsnSerGluSerAla-----TrpValCysValProAlaSerIleTyrCys 369
 QY 1021 AGGACAATGTTGAGGGCCATATACGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 Db 370 AsnAsnAlaLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArg 388
 QY 1081 CATCCATATGATGACCCGACT-----CCGCCAAGTTATTACAAACAAATTT 1125
 Db 389 GlyLysCysGluAspGluSerAsnLeuCysTyrLysGlyMetGlyTyrValSerGluTyr 408
 QY 1126 CTGGCAAAAGGACTCTGTCTATGACGCTATCGCGCTCAACATCAAC---TACACCCAGTCC 1182
 Db 409 LeuAsnLysProGluValArgAlaAlaValGlyAlaGluValAspGlyTyrAspSerCys 428
 QY 1183 AATAATGAGCTCTACTACGCTTTCAGCAACAGGAGCTTGTCTGGCCC-----AAC 1236
 Db 429 AsnPheAspIleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArg 448
 QY 1237 TTCATCGAAGACTCGAGGAGATCTTGTCTCTCCCGTCCGCTGCTCTCCCTCATCTAT--- 1293
 Db 449 LeuValProGlyIleLeuGlu-----GlnIleProVal-----LeuIleTyrAla 463
 QY 1294 GCGCAGCGCGATATACATCTGCAACTGGTTTCGGCGGTACGGCGTTCCTCGTGGGAAC 1353
 Db 464 GlyAspAlaAspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGlu 483
 QY 1354 TACTCCAAAGCGCCGAGTTCGAGACGCGAGGTACACGCCCTGGAAGTC----- 1404
 Db 484 TrpProGlyHisLysGluPheAlaAlaAlaProMetGluAspLeuLysIleValAspAsn 503
 QY 1405 -----AACGGCTCGAGTATGGGAAACTCGCGAGTATGGTAATTTCTCTCTCACTCGC 1458
 Db 504 GluHisThrGlyLysLysIleGlyGlnIleLysThrHisGlyAsnPheThrPheMetArg 523
 QY 1459 GTCTATGAGCGAGCGCATGAAGTCCCATCTACACCCCATCGCTCCCTGCAATTTGTT 1518
 Db 524 LeuTyrGlyGlyGlyHisMetValProMetAspGlnProGluAlaSerLeuGluPhePhe 543
 QY 1519 AACCGGACTATCTTCGCT---TGG 1539
 Db 544 AsnArgTrpLeuGlyGlyGluTrp 551

RESULT 4

O94152 PRELIMINARY; PRT; 541 AA.
 AC O94152;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Carboxypeptidase Y.
 GN CPY.
 OS Pichia angusta (Yeast) (Hansenula polymorpha).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 CX NCBI_TaxID=4905;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS4732;
 RX MEDLINE=99174803; PubMed=10077185;
 RA Bellu A.R., van der Klei I.J., Reehinger K.B., Yavuz M., Veenhuis M.,
 RA Kiehl J.A.K.W.;
 RT "Characterization of the Hansenula polymorpha CPY gene encoding
 RT carboxypeptidase Y";
 RL Yeast 15:181-189 (1999).
 DR EMBL; AF085063; AAC96121.1; -.
 DR HSSP; P00729; 1YSC.
 DR MEROPS; S10.001; -.

DR InterPro; IPR001563; Serine carboxypept.
 DR InterPro; IPR000379; Ser estrs site.
 DR Pfam; PF00450; serine carboxypept. 1.
 DR PRINTS; PR00724; CRBOXYPEPTASEC.
 DR ProDom; PD001189; Serine carboxpept; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 SQ SEQUENCE 541 AA; 60793 MW; A00CCB58526B1753 CRC64;

Alignment Scores:
 Pred. No.: 4,06e-34 Length: 541
 Score: 596.50 Matches: 164
 Percent Similarity: 46.14% Conservativeness: 63
 Best Local Similarity: 33.33% Mismatches: 184
 Query Match: 19.71% Indels: 81
 DB: 3 Gaps: 20

US-09-712-338-1 (1-1668) x O94152 (1-541)

QY	136	AACATGTCACATCCGGTACAAAG-----GAAACCGGGGAGAGGCGCTCTCGAGACT	189
Db	115	SerAsnTyrLysLeuArgValIlystysThrAspProGlyAlaLeuGlyLeuAspHisThr	134
QY	190	ACCCCGGGTGTCAAATCTACTCTCGATATGTCGACACCTCTCCCGAG---TCCCATACC	246
Db	135	-----ArgGlnTyrSerGlyTyrLeuAspValGluAspGluAspLysHisPhe	150
QY	247	TTCTTCTGGTCTTCGAAGCCAGACATACCCAGAACTGCACCTATCATCATGTGGTTG	306
Db	151	PheTyrTrpMetPheGluSerArgAsnAspProAlaAsnAspProValleLeuTrpLeu	170
QY	307	AATGTGGCCCTGGAAAGCATCTTTTGATCGGTCTCTTCCGAAGAGTTGGCCCTTGCAT	366
Db	171	AsnGlyGlyProGlyCysSerSerLeuThrGlyMetLeuPheGluLeuGlySerAlaSer	190
QY	367	GTCAATTCCGACTTTTGATGACTACATCACTACCTCACTCGTGAACAGAGTCTCCAAATT	426
Db	191	IleGlyProAspLeuLysProIleAsnAsnProTyrSerTrpAsnSerAsnAlaThrVal	210
QY	427	CTATTCTGTGCCAGCCATTGGGAGTCGGCTTTTCATATAGTGATACGGTTGATGGTCC	486
Db	211	IlePheLeuAspGlnProValAsnValGlyPheSerTyrSer-----	224
QY	487	ATTAACCTGTAACTGGGGTGTGCGAAATTGCGAGCTTTGCGAGGAGTTTCAGGGCCGGT	546
Db	224	-----	224
QY	547	CCAAACCATGTATGCCACTCTCTGATCGATACCTACCAATCTTCGGCAGAGCGCGCTGG	606
Db	225	-----SerLysSerValSerAsnThrValAlaAlaGlyLysAspValTyrAla	240
QY	607	ATCTGTCAGGAATTCCTTTAGTGGACTACTAGCTTGGACTCTAGGGTGCAGTCTAAGAC	666
Db	241	PheLeuGluLeuPheTyrGlnGlnPheProHisLeu-----LeuLysAsnAsp	256
QY	667	TTCACTCTATGGACGGAGAGCTATGGAGGCACTATGGTCTCGATTTCTCAATCATTTT	726
Db	257	PheHisIleAlaGlyGluSerTyrGlyHisTyrIleProValPheAlaSerGluIle	276
QY	727	TACGAGCAGAAATGAGAAATGGCCACCGTAGTGTATATGTGTGTTCAGCTTAATTCAAC	786
Db	277	LeuThrHisAlaAspArg-----SerPheAsnLeuThr	287
QY	787	TCCTCTGGGAATTTAATACGGCATCATCGACGAGCGGATCCAGGCGCCTTACTACCTGAA	846
Db	288	SerValLeuIleGlyAsnGlyLeuThrAspProLeuAsnGlnTyrProPheTyrGluArg	307
QY	847	TTCGCTGTGAACAATACCTACGGTATCAAG---CCTGTCAACGAGACCGTCTACAACCTAC	903
Db	308	MetAlaCysSerThrAspGlyGlyTyrGluProThrLeuAspGluSerGluCysGluGly	327
QY	904	ATAAGATTTTCCCAACCAAAATGCCAAATGGTTGCCAGGATTTGATTTCCACTGCAACAG	963
Db	328	Met-----LeuGluThrLeuProArg---CysLeuSerLeuIleGluSerCysTyrSer	344

DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.	CRC64;
SO	SEQUENCE 537 AA; 60270 MW; FC01E57D7582C941	

Alignment Scores:

Alignment Scores:			
Pred. No.:	4, 778-34	Length:	53
Score:	595.50	Matches:	165
Percent Similarity:	45.01%	Conservative:	65
Best Local Similarity:	32.29%	Mismatches:	196
Query Match:	19.67%	Indels:	85
DB:	3	Gaps:	20

US-09-712-338-1 (1-1668) x 014414 (1-537)

QY	91	CAGCTACCCAAAGAACCCCGGGTCAAGACTCTTAAACCGCAACAATGTCCACCATC	150
Db	92	LysilleThrLysLysProAlaAspPheThrAspPheAsnValLeuAsnGluLysPheSer	111
QY	151	CGGTACAG-----GAACCGGGGAGAGGGCGCTCGAGACTACC	192
Db	112	AsnTyrLysLeuArgValLysLysThrAspProGlyAlaLeuGlyLeuAspHisThr---	130
QY	193	CGGGTGTCAAATCCTACTCTGGATATGTGACACCTCTCCGAG-----TCCATACCTTC	249
Db	131	-----ArgLntyrSerGlyTyrLeuAspValGluAspGluAspLysHisPhePhe	147
QY	250	TTCGTGTTCTCGAGCCAGACATAACCCAGAACTGCACCTATCACATCTGTGTTCAAT	309
Db	148	TyrTrpMetPheGluSerArgAsnAspProValAsnAspProValLeuLeuTrpLeuAsn	167
QY	310	GGTGGCCCTGGAAGCGATCTTTTCATCGCTCTCTCGAAGAGTTGGGCCCTTGCATGTC	369
Db	168	GlyGlyProGlyCysSerSerLeuThrGlyMetLeuPheGluLeuGlySerAlaSerIle	187
QY	370	AAATCGACTTTTGATGACTACATCAACCTCACTCGTGAACGAGGTCTCCAAATTTACTA	429
Db	188	GlyProAspLeuLysProIleAsnAsnProTyrSerTrpAsnSerAsnAlaThrValile	207
QY	430	TTCTCTGCCACCCATTTGGAGTCCGGCTTTTCATATAGTGTATCGGTTGATGGGTCCATT	489
Db	208	PheLeuAspGlnProValAsnValGlyPheSerTyrSer-----	220
QY	490	AAACCTGTAACTGGGGTGCTGAAATTCGAGCTTTGAGAGTTTCAGGGCGGTACCCA	549
Db	220	-----	220
QY	550	ACCATGTATGCACCTCTGATGATACTACCAATCTTCGCGACAGAGCGCGTTGGAGATC	609
Db	221	-----SerLysSerValSerAsnThrValAlaAlaGlyLysAspValTyrAlaPhe	237
QY	610	CTGCAAGATCTCTTAGTGAGTACCTAGCTTGGACTTAGGGTGAGTCTTAAGGACTTC	669
Db	238	LeuGluLeuPheTyrGlnGlnPheProHisLeu-----LeuLysAsnAspPhe	253
QY	670	AGTCTATGACGGAGAGCTATCGAGGCACTATGGTCTCTGCATCTTCAATCATTTTTC	729
Db	254	HisIleAlaGlyLysSerTyrGlyGlyHisTyrIleProValPheAlaSerGluIleLeu	273
QY	730	GAGCAAAATCAGAAATTCACACGGTAGTGTAAATGGTGTAAATTTCACTCTCT	789
Db	274	ThrHisAlaAspArg-----SerPheAsnLeuThrSer	284
QY	790	CTGGGAATTTAAACGGCATCATCGACGAGCGATCCAGGCCCTTACTACTCACTTTC	849
Db	285	ValLeuIleGlyAsnGlyLeuThrAspProLeuAsnGlnTyrProPheTyrGluArgMet	304
QY	850	GCTGTGAACAAATACCTACGGTATCAAG-----GCTGTCAACGACGACCGCTCTCAACCTACATG	906
Db	305	AlaCysSerThrAspGlyGlyTyrGluProThrLeuAspGluSerGluCysGluGlyMet	324
QY	907	AAGTTTCCCAACCAATGCCAATGGTTGCCAGATTGTTTCCACTTGCACCAACAGACA	966
Db	325	-----LeuGluThrLeuProArg-----CysLeuSerIleGluSerTyrSerSer	341

QY	967	AACGGCACCGCAATTAGCTGACTACGGCCCTCTGGCGCGAAGCAACCAACATGTCGAGGAC	1021
Db	342	Gln-----SerValPheSerCysValProAlaSerIleTyrCysAsnAsn	356
QY	1027	AATGTTGAGGGGCCATACACTACGCCTTTGCTGTGGTGTGTGTATGATATTCGGCATCCA	1086
Db	357	AlaGlnLeuGlyProPheGlnLys---ThrGlyArgAsnValTyrAspValAlaArgLysMet	375
QY	1087	-----TATGATGACCCGACTCCGCCAAGTTATTACAACAAATTTCTG	1128
Db	376	CysGluGlyThrLeuCysTyrLysAsp-----MetGluTyrIleAspGlnTyrLeu	392
QY	1129	GCAAGGACTCTGTCATGAGCGCTATCGGCGTCAACATCAAC---TACACCCAGTCCCAAT	1185
Db	393	AsnGlnAspPheValLysGluLysValGlyAlaGluValAspThrTyrGluSerCysAsn	412
QY	1186	AATGACGCTACTACGCTTTCCACCAACAGGCGACTTTGCTGGCCC-----AAC	1236
Db	413	PheAspValAsnArgAsnPheLeuPheAlaGlyAspTrpMetLysProTyrHisLysAsn	432
QY	1237	TTCATCGAAGACCTCGAGGAGATCCTTGCTCTCCCGCTGGCGTGTCTCCCTCATCTAT---	1293
Db	433	VallIleAsnLeuLeuGluGln-----GlyLeuProVal-----LeuIleTyrAla	447
QY	1294	GCGGACGCCGATTACATCTGCAACTGGTTCGGCGGTGAGGCGGTTTCCTCGCTCGCAAC	1353
Db	448	GlyAspLysAspPheIleCysAsnTrpLeuGlyAsnGlnAlaTrpSerAsnGluLeuPro	467
QY	1354	TACTTCCCAAGCCCGCCAGTTCGGAAGCGCAGGG---TACAGCGCCCTGAAAGTCAACGGC	1410
Db	468	TrpSerGlyHisAspGluPheGluSerAlaGluLeuTyrAsnLeuThrLeuLysAspGly	487
QY	1411	GTCGAGTAGTGGGAAACTCGCGAGTATGTTAATTTCTCCTTCACTCGCGTCTATGAGGCA	1470
Db	488	ThrLysValGlyGluValLysAsnAlaGlyLysPheThrPheAlaArgMetPheAspGly	507
QY	1471	GGCCATGAAGTCCCATACTACAGCCATCGCTCCCTCGCAATGCTTTTAACGGGACTATC	1530
Db	508	GlyHisMetValProTyrAspGlnProGluSerSerLeuAlaMetValAsnArgTrpIle	527
QY	1531	TTCGGTGGTATTCGACAGGGCCAGAGAAG	1563
Db	528	AlaGly---AspTyrSerLeuGlyThrLysLys	537
RESULT 6			
Q8VWQ0		PRELIMINARY; PRT; 507 AA.	
ID	Q8VWQ0		
AC	Q8VWQ0;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Putative serine carboxypeptidase.		
GN	SCP.		
OS	Gossypium hirsutum (Upland cotton).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Malvales; Malvaceae; Malvoideae; Gossypium.		
OX	NCBI_TaxID=3635;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAT=cn, Xu-142; TISSUE=Fiber;		
RA	Li C.-H., Zhu Y.-Q., Meng Y.-L., Wang J.-W., Xu K.-X., Zhang T.-Z.,		
RA	Chen X.-Y.,		
RT	"Isolation of Genes Preferentially Expressed in Cotton Fiber Cells by		
RT	cDNA Array and RT-PCR."		
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY072822; AAL67992.1; -		
DR	MEOS; S10.009; -		
DR	InterPro; IPR001563; Serine_carbpept.		
DR	InterPro; IPR000379; Ser_estrs_site.		
DR	Pfam; PF00450; serine_carbpept; 1.		
DR	PRINTS; PD00724; CRBOXYPTASEC.		
DR	ProDom; PD001189; Serine_carbpept; 1.		

08VW00
RESULT

```

ID   Q8VWQ0 PRELIMINARY;      PERT;      507 AA.
AC   Q8VWQ0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative serine carboxypeptidase.
GN   SCP.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OC   NCBI_TaxID=3635;
[1]
RN   RP
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Xu-142; TISSUE=Fiber;
RA   Li C.-H., Zhu Y.-Q., Meng Y.-L., Wang J.-W., Xu K.-X., Zhang T.-Z.,
RA   Chen X.-Y.;
RT   "Isolation of Genes Preferentially Expressed in Cotton Fiber Cells by
RT   cDNA Array and RT-PCR.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY072822; AAL67992.1; -.
DR   MEROPS; S10.009; -.
DR   InterPro; IPR001563; Serine_carbpept.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00450; serine_carbpept_1.
DR   PRINTS; PR00724; CRBOX1PTASEC.
DR   ProDom; PD001189; Serine_carbpept; 1.

```

```

DR PROSITE, PS00560; CARBOXYPEPT SER HIS; 1.
KW PROSITE, PS00131; CARBOXYPEPT_SER_SER; 1.
DR Carboxypeptidase.
SQ SEQUENCE 507 AA; 56594 MW; 6A0D283D9C5E7713 CRC64;

Alignment Scores:
Pred. No.: 9.66e-26 Length: 507
Score: 479.00 Matches: 149
Percent Similarity: 43.93% Conservative: 90
Best Local Similarity: 27.39% Mismatches: 205
Query Match: 15.82% Indels: 100
DB: 10 Gaps: 21

US-09-712-338-1 (1-1668) x Q8VWQ0 (1-507)

QY 37 GTTGACAGCTGGGCGCTTCAGAGGAGTACACCGGCGTCCGTGCGTAGACAGCACTA 96
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 27 IleThrAlaAspLeuArgLeuThrAspSerPheProSerValHisAlaLeuLysLeu 46
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 97 CCCAAGAAC-----CCACCGGGGTCAAGACTCTTACACCGCAAAACAATGTC 144
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 47 IleArgGluLeuAsnLeuPheProLysGlyGluValAsnValValAspGluHisArgVal 66
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 145 ACCATC-----CGTACAAAGAAACCCGGGGCAGAGGGC 177
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 67 SerLeuProGluGlyProLysLeuValGluLysArgPheLysPheProThrLeuGlu--- 85
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 178 GTCTGCGAGACTACCCGGGTGTCAAATCCTAC-----TCTGGATAT 219
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 86 -----ValProGlyGlyValSerPheGluAspLeuGlyHisAlaGlyTyr 101
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 220 -----GTCGACACCTCTCCGAGTCCCATACCTCTCTGTTCTTCGAGAGCCAGACAT 273
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 102 TyrLysLeuProAsnSerHisAspAlaArgMetPheTyrPhePheGluSerArg--- 120
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 274 AACCCAGAACTGACTATCAATGTGTGTTGAATGGTGGCCCTCGAAGCGATCTTTTG 333
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 AsnSerLysLysAspProValValIleTyrLeuThrGlyGlyProGlyCysSerSerGlu 140
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 334 ATCGGTCTCTTCGAAGAGTTGGCCCTTGCATGTCAATTCGACTTTTGTAGTACATC 393
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 141 LeuAlaLeuPheTyrGluAsnGlyProPheThrIleAlaAspAsnMetSerLeuValTyr 160
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 394 AACCCCTCACTGTGGAACGAGGTCTCCAAATTACTATTCTGTCGCCAGCCATGGGAGTC 453
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 161 AsnGluTyrGlyTyrAspLysAlaSerAsnLeuLeuTyrValAspGlnProIleGlyThr 180
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 454 GGTCTTTCATATAGTATACGTTGATGGTCCCATTAACCCCTGTAACCTGGGTGTCGAA 513
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 GlyPheSerTyrSer----- 185
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 514 AATTGAGCTTTTCAGAGTTACAGGCGGTACCCCAACCATTCATGCCACTCTGATCGAT 573
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 186 ---SerAspGlnArgAspIleArgHisAsnGluAspGluValSerAsnAspLeuTyrAsp 204
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 574 ACTACCAATCTTCCGCGAGAGCGGCTGGAGATCCTGCAAGGATTCCTTAGTGACTA 633
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 205 -----PheLeuGlnAlaPhePheAlaGluHis 213
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 634 CTTAGCTTGAGCTCTAGGTGCGAGTCTAAGGACTTCAGTCTATGACGAGAGCTATGGA 693
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 214 ProGluPhe-----AlaLysAsnAspPhePheIleThrGlyGluSerTyrAla 229
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 694 GGGCACTATGGTCTCANTCTTCATCATTTTACGACAGATCAGAGAATTCGCCAAC 753
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 230 GlyHisTyrIleProAlaPheAlaAlaArgValHisArgLysLys----- 245
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 754 GGTAGTGTAAATGGTGTTCAGCTTAAATTCAACTCTCTGGGAATTAAACGSCATCATC 813
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 246 ---AlaLysGluGlyIleHisIleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThr 264
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 814 GAGAGGCGATCCAGGCCCTTACTACCTGTAATTCGCTGTGAACAATACCTACGCTATC 873
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 7

```

Q9XH61
ID Q9XH61 PRELIMINARY; PRT; 501 AA.
AC Q9XH61;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Serine carboxypeptidase.
OS Matricaria chamomilla.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Anthemideae; Matricaria.
OX NCBI_TaxID=98504;
RN (1)
RP SEQUENCE FROM N.A.
RA Kohchi C., Yasuda H., Hirata T.;
RT "Isolation of a cDNA encoding for a carboxypeptidase, having leucine
RT zipper structure at the N-terminal region, from the cultured shoot

```

QY	607	ATCCTGCAAGGATTCCTTACTGAGCACTACCTAGCTTGGACTCTAGGCTGCAGTCTAAGGAC	666
DB	198	PhcLeuGlnAlaPheThrGluHisProGluPhe-----ValAspAsnAsp	213
QY	667	TTCAAGTCTATGGACGGAGAGACTATGAGGGCACTATGGTCTCGCATTTCTCAATCATTTT	726
DB	214	PhcTyrIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaIleAlaAlaArgVal	233
QY	727	TACGACGAGAATGAGAGAATGCGACGGTAGTGTTAATGGTGTTCAGCTTAATTTCAAC	786
DB	234	HisGlnGlyAsnLys-----AlaLysGluGlyIleHisIleAsnLeuLys	248
QY	787	TCTCTGGGAATTATTAAACGCATCATCGACGGCGCATCCAGGCCCTTACTACCCCTGAA	846
DB	249	GlyPheGlyIleGlyAsnGlyLeuThrAspProAlaIleGlnTyrGlnAlaTyrThrAsp	268
QY	847	TTTCGTGTGAACAACTACTACGGTATCAAGGCTGTCAACGAGACCGCTTCAACAATCATG	906
DB	269	TyrAlaLeuAsp-----MetGlyLeuIleLysGluSerGlnTyr-----	281
QY	907	AAGTTTCGCAACCAATGCCAATGGTTGGCAGGATTTGATTTCCACCTGCAACAGACA	966
DB	282	LysArgIleAsnLeuIleValProLeuCysGluAlaAlaIleLysLeuCysGlyThrAsp	301
QY	967	AACCGC---ACCGCATTAGTGTACTACGCCCTCTGCGCCGAAGCCACCAATGTGCAGG	1023
DB	302	GlyThrValSerCysMetAlaTyrValValCys-----	313
QY	1024	GACAATGTGAGGGCCATCTACGCTTGTCTGCTGCTGCT---GGTGTGTATGATATTCGG	1080
DB	314	AsnThrIlePheSerSerIleLeuSerIleAlaGlyAsnIleAsnTyrTyrAspIleArg	333
QY	1081	CATCCA-----TATGATGACCCGACTCCGCCAAGTATTATCAACAAA	1122
DB	334	LysGluCysValThrSerMetCysTyrAspPhe-----SerAspMetGluThr	349
QY	1123	TTTCTGCGAAAGGACTCTGTCATGAGACGCTATCGGGTCT---AACATCACTACACCCAG	1179
DB	350	LeuLeuAsnLysLysSerValArgGlnAlaLeuGlyValGlyAspIleGluPheValSer	369
QY	1180	TCCAAATAATGACGTCTACTACGCTTCCAGCAACAGCGACTTGTCTGGCCCACTTC	1233
DB	370	CysSerThrThrValTyrThrAlaMetLeuValAsp-----LeuMetArgAsnLeu	386
QY	1240	ATCGAAGACCTCGAGGAGACTCTGCTCTCCCGCTGGCTGCTCCCTCATCTATGGCGAC	1299
DB	387	GluAlaGlyIleProGluLeuLeuGluAspGlyIleLysMetLeuValTyrAlaGlyGlu	406
QY	1300	GCGGATTCATCTGCAACTGGTGTGGCGGTGAGGCCGTTTCCCTGCTGCGAAGTACTCC	1355
DB	407	TyrAspValIleCysAsnTrpLeuGlyAsnSerArgTrpValHisAlaMetGluTrpAsn	426
QY	1360	CAAGCGCCCAAGTTCGGAAGCGAGGTACACGCCCCCTGAAAGTCAAAGCGGTCCAGTAT	1411
DB	427	GlyLysGluGlnPheAsnAlaLeuSerGluAlaProPheGluValAspGlySerGluAla	446
QY	1420	GGGGAATTCGCGAGTAGTGGTAATTTCTCTTCACTCGCGTCTATGAGCGAGCCATGAA	1471
DB	447	GlyLeuLeuLysSerTyrGlyProLeuSerPheLeuLysValHisAspAlaGlyHisMet	466
QY	1480	GTCCCATACTACCGCCCATCGCTCCCTCGCAATGTTTAAACCGGACTATCTCGGT	1536
DB	467	ValProMetAspGlnProLysAlaAlaLeuGluMetLeuLysArgTrpMetAspGly	485
RESULT 8			
Q8VNT1	PRELIMINARY; PRT; 429 AA.		
AC	Q8VNT1		
DT	01-WAR-2003 (TrEMBLrel. 23, Created)		
DT	01-WAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-WAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Carboxypeptidase C cbp31.		
GN	P0434A03.7.		

OS *Oryza sativa* (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriharoidae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GAJ) genomic DNA, chromosome 7, PAC
RT clone: P043A03."
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004299; BC45113.1; -.
SQ SEQUENCE 429 AA; 47790 MW; 1F2F64F236475BB4 CRC64;

Alignment Scores:
Pred. No.: 2,45e-25 Length: 429
Score: 473.00 Matches: 130
Percent Similarity: 45.4% Conservatives: 77
Best Local Similarity: 28.5% Mismatches: 178
Query Match: 15.6% Indels: 70
DB: 10 Gaps: 16

US-09-712-338-1 (1-1668) x Q8GV11 (1-429)

QY	208	TACTCTGATAT-----GTCGACACTCTCCGAGTCCCACTCTCTTCTGTTCTTC	261
Db	17	HisAlaGlyTyrTyArgLeuProAsnThrHisAspAlaArgLeuPheTyrPhePhe	36
QY	262	GAAGCAGACATAACCCA--GAACCTGCACCTATCACATCTGTTGTTGAATGGTGGCCCT	318
Db	37	GlueSerArgGlySerGlyGluAspAspProValValIleTrpLeuThrGlyGlyPro	56
QY	319	GGAAGCGAATCTTTGATCGGTCTCTTCGAAGAGTTGGGCCCTTGCCATGTCATTCGACT	378
Db	57	GlyCysSerSerGluLeuAlaLeuPheTyrGluAsnGlyProPheHisIleAlaAspAsn	76
QY	379	TTTGATGACTACATCAACCTCCTCCTGTTGGAACGAGGTCTCCAAATTAATCTCTGCTCC	438
Db	77	MetSerLeuValTrpAsnAspPheGlyTrpAspGlnGluSerAsnLeuIleTyrValaAsp	96
QY	439	CAGCCATTGGGAGTCGGGTTTCATATAGTGATACGGTTGATGGGTCCCATTAACCTGTGA	498
Db	97	GlnProThrGlyThrGlyPheSerTyrSerSerAsnProArgAspThrArgHisAspGlu	116
QY	499	ACTGGGTGCTGAAAATTCAGGCTTTCAGGAGTTCAGGCGCGGTACCCACCACTTGAT	558
Db	117	AlaGly--ValSerAsnAspLeuTyrAla-----	125
QY	559	GCCACTGTGATCGATACCAATCTTCCGCGAGGCGGCTTGGGAGATCTGCAAGGA	618
Db	126	-----PheLeuGlnAla 129	
QY	619	TTCTTAGTGGACTACCTAGTCTTGAGCTCTAGGTGCAGTCTAAGGACTCTCAGTCTATGG	678
Db	130	PhePheThrGluHisProAsnPhe-----AlaLysAsnAspPheTyrIleThr 145	
QY	679	ACGAGAGCTATGAGGGCACTATGTCCTGCATCTTCAATCAATTTTACGAGCAGAAAT	738
Db	146	GlyGluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValTyrLysGlyAsn 165	
QY	739	GAGAGAATTGCCAAGGTAGTGGTTAATGGTCTTCAGCTTAATTTCACTCTCTCTGGGAAT	798
Db	166	Lys-----AsnSerGluGlyIleHisIleAsnLeuLysGlyPheAlaIle 180	
QY	799	ATTAACGGCATATPCGAGGGCGGATCCAGGCCCTTACTACCTGAATTCGCTGTGAAC	858
Db	181	GlyAsnGlyLeuThrAspProAlaIleGlnTyrLysAlaTyrThrAspTyrSerLeuAsp 200	
QY	859	AATACCTACGGTATCAAGGCTGTCAACGAGACCGCTTACAACTACATCAAGTTTGCCAAAC	918
Db	201	-----MetGlyLeuIleThrLysSerGlnPheAsnArgIle-----Asn 213	

DR PRINTS; PR00724; CRBOXYPTASEC.

DR PRODOM; PD001189; Serine carboxypept; 1.

DR PROSITE; PS00560; CARBOXYPEPT SER HIS; 1.

DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.

KW Carboxypeptidase.

SQ SEQUENCE 505 AA; 56543 MW; 6E74CD351090B099 CRC64;

Alignment Scores:

Pred. No.:	7.52e-25	Length:	505
Score:	466.50	Matches:	132
Percent Similarity:	44.42%	Conservative:	75
Best Local Similarity:	28.33%	Mismatches:	178
Query Match:	15.41%	Indels:	81
DB:	10	Gaps:	15

US-09-712-338-1 (1-1668) x Q9FPB0 (1-505)

QY	193	CGGGTCTTCAATCC-----TACTCTGGATAT-----	GTGACACCTCTCCCGAG	237
Db	91	ProSerValGlnGluPheGlyHisTyrAlaGlyTyrSerLeuProHisSerLysSer	110	
QY	238	TCCCATACCTCTCTCTGGTTCTTCGAAGCCAGACATACCCAGAACTGCACCTATCACA	297	
Db	111	AlaLysMetPheTyrPhePhePheGluSerArg---AsnLysThrThrAspProValVal	129	
QY	298	TTGTGGTTGAATGGTGGCCCTGGAGCGGATCTTTGATCGGTCTCTTCGAAGAGTTGGGC	357	
Db	130	IleTrpLeuThrGlyGlyProGlyCysSerSerValAlaMetPheTyrGluAsnGly	149	
QY	358	CTTTGCCATTCGAATTCGACTTTTGTGATGATCATCATCAACCCCTCACTCGTGGACAGGTC	417	
Db	150	ProPheLysIleSerLysAspLeuSerLeuTyrTrpAsnAspPheGlyTrpAspLysVal	169	
QY	418	TCCAATTACTATTCCTGTCCAGCAATGGAGTCGGCTTTTCATATGATGATCGGTT	477	
Db	170	SerAsnIleIleTyrValAspGlnProValGlyThrGlyPheSerTyrThrSerAspGlu	189	
QY	478	GATGGTCCATTACCCCTGTAACTCGGGTGTGCGAAATTCGAGCTTGCGAGGATTCAG	537	
Db	190	SerAspIleArgAsnAspGluAspGlyVal-----	199	
QY	538	GGCCGGTACCCCAACCAATGTATGCCACTCTGATCGATACTACCAATCTGCCGACAGGCG	597	
Db	200	-----SerAsnAsp	202	
QY	598	GCTTGGGAGATCTCGCAAGGATTCCTTAGTGGACTACTAGCTTGGACTTAGAGTGCAG	657	
Db	203	LeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro-----LysPheVal	218	
QY	658	TCTAAGGACTTCAGTCTATGACGCGAGAGCTATGAGGCGCACTATGCTCGTCAATCTTC	717	
Db	219	LysAsnAspPhePheIleThrGlyCysTyrAlaGlyHisTyrIleProAlaLeuAla	238	
QY	718	AATCAATTTTTCAGCAGCAATACGAGAAATTCGCCAGGTAGTGTAAATGTTGTTCAGCTT	777	
Db	239	SerArgValHisSerGlyAsnLysLys-----LysGluGlyIleProIle	253	
QY	778	AATTTCAACTCTCTGGGAATTTATTAACGGCATCATCGACGGCGATCCAGGCCCTTAC	837	
Db	254	AsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAsnProGluIleGlnTyrGlyAla	273	
QY	838	TACCCCTGAATTCGCTGCAACAATCACTACGGTATCAAGGCTGTCACACGACGCGTCTAC	897	
Db	274	TyrGlyAspTyrAlaLeuGln-----MetLysLeuIleSerGluSerAspHis	289	
QY	898	AACTACTACGAATTTGGCCAAACCAATGCCAAATGGTTCGCCAGGATTTGATTTCCACTGC	957	
Db	290	GluSerLeuLys-----GlnAspTyrValGluCysGln	300	
QY	958	AAACAGACAAACCGCAGCATTTAGCTGACTACGCCCTCTCGCGCGAGGCCACCAACATG	1017	
Db	301	AsnIleThrLysCysSerLeuGlyGlyGlyLeuValCysAspSerAlaValGluVal	320	

281 Db -----MetGlyValIleLysLysSerAspTyrAsnArgIle-----AsnLysLeu 295

QY 925 CCAATGGTTGCCAGGATTGATTTCACCTGCACACAGACAACCCGC---ACGCAATTA 981

Db 296 ValProValCysGluMetAlaIleLysLeuCysGlyThrAspGlyThrIleSerCysMet 315

QY 982 GCTGACTACGCCCTCTCGCCGAGGCCACCAACATGTGCAGGACAAATGTAGGGGCCA 1041

Db 316 AlaSerTyrPheValCys-----AsnAlaIlePheThrGly 327

QY 1042 TACTACGCCTTTGCTGCT---CGTGTGTGTATATATTCGGCATCCATATGATGACCCG 1098

Db 328 IleMetAlaLeuAlaGlyAspThrAsnTyrTyrAspIleArgThrLysCysGluGlySer 347

QY 1099 ACTCGGCCAAGTTATTACAC-----AAATTTCGGCAAAGAACTCTGTCATGGACGCT 1152

Db 348 LeuCysTyrAspPheSerAsnMetGluThrPheLeuAsnGlnGluSerValaArgAspAla 367

QY 1153 ATCGGCGTC---AACATCAACTACACCCAGTCCCAATATGACGCTTACTACGCTTCCACG 1209

Db 368 LeuGlyValGlySerIleAspPheValSerCysSerProThrValTyrGlnAlaMetLeu 387

QY 1210 CAACAGCGCACTTTGCTGCGCCAACTTCATCGAA---GACCTCGAGGAGATCTTGTCT 1266

Db 388 ValAsp-----TpmMetArgAsnLeuGluValGlyIleProAlaLeuLeuGlu 403

QY 1267 CTCCTCGTGGTCTCCCTCATCTATGCGCGAGCGATTACATCTGCAACTGTTTCGGC 1326

Db 404 AspGlyValLysLeuValTyrAlaGlyGluTyrAspLeuIleCysAsnTrpLeuGly 423

QY 1327 GGTGAGCGCGTTTCCCTCGCTCGGACTACTCCCAAGCGCCAGTCCCAAGCGCAGG 1386

Db 424 AsnSerArgTrpValHisAlaMetGluTrpSerGlyGlnLysGluPheValAlaSerPro 443

QY 1387 TACACGCCCTGAAAGCTCAACGCGCTCGAGTAGGGGAACTCGCGAGTAGTGTAATTC 1446

Db 444 GluValProPheValValAspGlySerGluAlaGlyValLeuArgThrHisGlyProLeu 463

QY 1447 TCCTTCACTCGCTCTATGAGCGAGCGCATGAGTCCCATCTACGAGCCCATCGCTCC 1506

Db 464 GlyPheLeuLysValHisAspAlaGlyHisMetValProMetAspGlnProLysAlaAla 483

QY 1507 CTGCAATTGTTTAAACCGGCACTCTTCGGTGG-----GATATCGCAGAGGCCAGAAG 1560

Db 484 LeuGluMetLeuLysArgTrpThrLysGlyThrLeuSerGluAlaAlaAspSerGluLys 503

QY 1561 AAGATC 1566

Db 504 LeuVal 505

RESULT 11

Q9LXH4 PRELIMINARY; PRT; 510 AA.

ID Q9LXH4 AC

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Carboxypeptidase precursor-like protein (AT3g45010/F14D17_80).

GN F14D17_80.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Eumatiophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Jordan N., Bancart S., Wiedelmann R., Voss H., Unseld M., Mewes H.W.,

RA Rudd S., Lencze K., Mayer K.F.X., Quetier P., Salaroubat M.;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases

[3] SEQUENCE FROM N.A.
 RA Koesema E., Chen H., Cheuk R., Kim C.J., Meyers M.C., Shinn P.,
 RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 [4]
 SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis ORF clones";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; ALJ53992; CAB89316.1; -;
 DR EMBL; AY050427; AAK91443.1; -;
 DR EMBL; AY093993; AAM16254.1; -;
 DR HSSP; P00729; 1YSC.
 DR MEROPS; S10.009; -;
 DR InterPro; IPR001563; Serine carboxypept.
 DR InterPro; IPR000379; Ser esterase site.
 DR Pfam; PF00450; serine carboxypept; 1.
 DR PRINTS; PR00724; CRBOXYPASEC.
 DR ProDom; PD001189; Serine carboxypept; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 SQ SEQUENCE 510 AA; 56896 MW; 134BC30F4E64BF0A CRC64;
 Alignment Scores:
 Pred. No.: 6,93e-23 Length: 510
 Score: 439.00 Matches: 130
 Percent Similarity: 42.43% Conservative: 83
 Best Local Similarity: 25.90% Mismatches: 191
 Query Match: 14.50% Indels: 98
 DB: 10 Gaps: 16
 US-09-712-338-1 (1-1668) x Q9LXH4 (1-510)
 QY 70 CCGGGTCCGTGGTAGAGACAGCTACCCAGAACCCCGGGGTCAAGACTCTTACA 129
 DB 82 ProAlaAlaValAspArg 88
 QY 130 ACCGCAACAATGTCACCATCGGTACAGGAACCCCGGGGCGGTGGAGACT 189
 DB 89 -----GlySerglygly----- 92
 QY 190 ACCCGGGTGTCAATCTCTAC-----TCTGGATAT-----GTGACACTCTCCC 234
 DB 93 SerProSerValGlnAspPheGlyHisAlaGlyTyrTyrLysLeuProAsnSerLys 112
 QY 235 GAGTCCCATACCTCTCTCTGGTCTTCGAGCCGACATACCCAGAACTGCACTATC 294
 DB 113 ALAAlaArgMetPheTyrPhePheGluSerArgThrAsn-----LysAlaAspProval 131
 QY 295 ACATTTGGTGAATGGTGGCCCTGGAAGCGANTCTTTGATCGGTCTCTCTCGAAGATTG 354
 DB 132 valilelrpLeuThrGlyGlyProGlyCysSerSerGluLeuAlaLeuPheTyrGluAsn 151
 QY 355 GGCCCTTGCCATGCAATTCGACTTTTGATGACTACATCAACCTCTACTGTGGAACGAG 414
 DB 152 GlyProPheThrValSerAsnAsnSerSerLeuSerTrpAsnGluPheGlyTrpAspLys 171
 QY 415 GTCTCCCAATTACTATTCTCTGCTCCAGCAATGGAGTGGCGCTTTTCATATAGTGATACG 474

DB 172 AlaSerAsnLeuIleTyrValAspGlnProValGlyThrGlyPheSerTyrThrSerAsp 191
 QY 475 GTTGATGGGTCCATTAACCTGTAACTGGGTGCTGCGAAATTCGAGCTTTCGAGGAGTT 534
 DB 192 GlnSerAspLeuArgHisAspGluAspGlyVal----- 202
 QY 535 CAGGGCCGGTACCCCAACCATTTGATCCACTCTGTATCGATACCTACCAATCTTCCGCGAG 594
 DB 203 -----SerAsn 204
 QY 595 GCGGCTTGGGAGATCTCGAAGGATTCCTTAGTGAGCTACCTAGCTTGGACTCTAGGGTG 654
 DB 205 AspLeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro-----GlnPhe 220
 QY 655 CAGTCTAAGGACTTTCAGTCTATGAGCGGAGAGCTATGAGGCGACTATGCTCTGCTGCTTC 714
 DB 221 ValLysAsnAspPheTyrIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaLeu 240
 QY 715 TTCATCATTTTACAGCAGAGATGAGAGAATGCCAAGGTAGTGTATATGGTGTTCAG 774
 DB 241 AlaSerArgValHisArgGlyAsnLys-----AsnLysGluGlyThrHis 255
 QY 775 CTTAATTTCAACTCTCTGGGAATTATTAAACGGCATCATCGAGCGGATCCAGGCCCT 834
 DB 256 IleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAsnProGluIleGlnTyrGly 275
 QY 835 TACTACCTGAATTCGCTGTGAACATACCTACGGTATCAAGGCTCTCAACGAGCGGTC 894
 DB 276 AlaTyrAlaAspTyrAlaLeuAspMetAsnLeuIleThrGlnSerAspHisAspLeu 295
 QY 895 TACAATCATCAAGTTTCCCAACCAATGCCAAATGGTGTGCCAGATTTGATTTCCACC 954
 DB 296 AsnArgTyr-----TyrAlaThr 301
 QY 955 TGCACACAGACAAACCGCAGCATTAGCTGACTAGCC-----CTCTCGCGCGAA 1005
 DB 302 CysGlnGlnSerIleLysGluCysSerAlaAspGlyGlyGluAspAlaCysAlaSer 321
 QY 1006 GCCACCAACATGTGCGAGGACAAATGTTGAGGGGCCATCTACGCCCTTGTGTCGTGT 1062
 DB 322 SerTyrThrValCys-----AsnAsnIlePheGlnLysIleMetAspIleAlaGlyAsnVal 340
 QY 1063 GGTGTGTATGATATTCGGCATCCATATGATGACCCGACTCCGCAAGTATTATCAAC 1119
 DB 341 AsnTyrTyrAspValArgLysGlnCysGluGlySerLeuCysTyrAspPheSerAsnMet 360
 QY 1120 ---AAATTTCTGCAAGGACTCTGTCTGAGCGCTATCGCGCTC---AACATCAACTAC 1173
 DB 361 GluAsnPheLeuAsnGlnLysSerValArgLysAlaLeuGlyValGlyAspIleGluPhe 380
 QY 1174 ACCAGTCCCAATTAATGAGTCTACTACGCTTTTCAGCAACACAGGCGACTTTGTCTGCC 1233
 DB 381 ValSerCysSerThrAlaValTyrGluAlaMetGlnMetAsp-----TrpMet 396
 QY 1234 AACTTCATCGAA---GACCTCGAGGAGATCTTGTCTCTCCCGGTGGTGTCTCCCTCATC 1290
 DB 397 ArgAsnLeuGluValGlyIleProAlaLeuLeuGlnAspGlyIleLysLeuValTyr 416
 QY 1291 TATGGCGACCCCGCATCTGCACTGCTGCTGCGGTTCAGGCGGTTCAGGCGGTTCCTCGCT 1350
 DB 417 AlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerLysTrpValHisGluMet 436
 QY 1351 AACTACTCCCAAGCCCGCGAGTCCGAGCGAGGTACAGCGCCCTGAAGTCAACGGC 1410
 DB 437 GluTrpSerGlyGlnLysGluPheValAlaAlaThrValProPheHisValAspAsn 456
 QY 1411 GTCAGTATGGGAACTCGCGAGTATGTAATTTCTCTTCACTCGCGCTATGAGGCA 1470
 DB 457 LysGluAlaGlyLeuMetLysAsnTyrGlySerLeuThrPheLeuLysValHisAspAla 476
 QY 1471 GGCCATGAAGTCCCATACTACCGCCCATCGCTCTCCCTCAATGTTTAAACCGGACTATC 1530
 DB 1530

```

Db 477 GlyHisMetValProMetAspGlnProValalaalaLeuGlnMetLeuGlnAsnTirpMet 496
QY 1531 TTCGGT 1536
Db 497 GlnGly 498
RESULT 12
Q932C3 PRELIMINARY; PRT; 510 AA.
AC Q932C3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ATG45010/F14D17 80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY057639; AAL15270.1; -.
DR InterPro; IPR001563; Serine carboxypeptidase.
DR InterPro; IPR000379; Ser esterase site.
DR Pfam; PF00450; serine carboxypeptidase; 1.
DR PRINTS; PR00724; CROXYPTASEC.
DR ProDom; PD001189; Serine carboxypeptidase; 1.
DR PROSITE; PS00560; CARBOXYPEPT SER HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT SER SER; 1.
SQ SEQUENCE 510 AA; 56837 MW; 704BC00C4E64BF08 CRC64;

Alignment Scores:
Pred. No.: 8,16e-23 Length: 510
Score: 438.00 Matches: 130
Percent Similarity: 42.23% Conservative: 82
Best Local Similarity: 25.90% Mismatches: 192
Query Match: 14.47% Indels: 98
DB: 10 Gaps: 16

US-09-712-338-1 (1-1668) x Q932C3 (1-510)
QY 70 CCGGCGTCCTCGGTAGACAGACGACTACCCAGAACCCCGGGGTCAAGACTCTTACA 129
Db 82 ProAlaValAspArg 88
QY 130 ACCGCAAAATGTCAACATCCGCTACAGGAACCCCGGGGAGGGCGCTCGAGACT 189
Db 89 -----GlySerGly 92
QY 190 ACCCGGGTCTCAATCCTAC-----TCTGGATAT-----GTCGACACTCTCCC 234
Db 93 SerProSerValGlnAspPheGlyHisAlaGlyTyrTyrLysLeuProAsnSerLys 112
QY 235 GAGTCCCATACCTTCTCTGGTCTTCGAGCCAGACATACCCAGAACTGCACCTATC 294
Db 113 AlaAlaArgMetPheTyrPhePheGluSerArgThrAsn-----LysAlaAspProVal 131
QY 295 ACATTGTGGTGAATGGTGGCCCTCGAAGCGATTCTTTGATCGGTCTCTTCGAAGAGTTG 354
Db 132 ValIleTrpLeuThrGlyGlyProGlyCysSerSerGluLeuAlaLeuPheTyrGluAsn 151
QY 355 GGCCTTGGCATGTCAATTCGACTTTTGTATGATCATCAACCCCTCACTCGTGGACGAG 414
Db 414 ----- 476

```

```

Db 152 GlyProPheThrValSerAsnAsnSerSerLeuSerTirpAsnGluPheGlyTirpAspLys 171
QY 415 GTCTCCAATTACTATCTCTGCCAGCCATGTTGGAGTCGGCTTTTTCATATAGTATACG 474
Db 172 AlaSerAsnLeuIleTyrValAspGlnProValGlyThrGlyPheSerTyrThrSerAsp 191
QY 475 GTTGATGGTCCATTAAACCCCTGTAACTGGGGTCTCGAAAATTCGAGCTTTCGAGGAGTT 534
Db 192 GlnSerAspLeuArgHisAspGluAspGlyVal-----SerAsn 202
QY 535 CAGGCGCGGTACCCCAACCAATTGATGCCACTCTGATCGATACCTACCAATCTTGGCCGAGAG 594
Db 203 ----- 204
QY 595 GCCCTTGGGAGATCTCTGCAAGGATCTTCTAGTGGACTAAGTCTTGGACTTAGGCTG 654
Db 205 AspLeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro-----GlnPhe 220
QY 655 CAGTCTAAGGACTTCACTATGACGAGGAGCTATGAGGGCACTATGGTCTTCGCTTC 714
Db 221 ValLysAsnAspPheTyrIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaLeu 240
QY 715 TTCAATCATTTTACGAGCAGATGAGAGAATTCGCCAACGGTACTGTAAATGGTGTTCAG 774
Db 241 AlaSerArgValHisArgGlyAsnLys-----AsnLysGluGlyThrHis 255
QY 775 CTTAATTTCACTCTCTGGGAATTTATTAACGGCATCATCGAGCGGATCCAGGCCCT 834
Db 256 IleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAsnProGluIleGlnTyrGly 275
QY 835 TACTACCTCGAATTCGCTGTGAACAACTACCTACGGTATCAAGCTGTCAAGGACCGTC 894
Db 276 AlaTyrAlaAspTyrAlaLeuAspMetAsnLeuIleThrGlnSerAspHisAspAsnLeu 295
QY 895 TACAACCTACATGAAGTTTGCCAAACCAATGCTTCCAGGATTTGATTTCCACC 954
Db 296 AsnArgTyr-----TyrAlaThr 301
QY 955 TGCAACAGACAAACCGCAGCATTTAGCTAGCTACGCC-----CTCTGCGCGAA 1005
Db 302 CysGlnGlnSerIleLysGlyCysSerAlaAspGlyGlyGluGlyAspAlaCysAlaSer 321
QY 1006 GCCACCAACATGTGCGAGGACAACTGTGAGGGCCATCTACGGCTTGTGGTGGT--- 1062
Db 322 SerTyrThrValCys---AsnAsnIlePheGlnLysIleMetAspIleAlaGlyAsnVal 340
QY 1063 GGTGTGTATGATTTGGCATCCATATGATGACCCGACTCCGCAAGTTTATTAACAC--- 1119
Db 341 AsnTyrTyrAspValArgLysGlnCysGluGlySerLeuCysTyrAspPheSerAsnMet 360
QY 1120 ---AAATTTCTGCAAGGACTCTGTATGAGCGCTATCGCGCTC---AACATCAACTAC 1173
Db 361 GluAsnPheLeuAsnGlnLysSerValArgLysAlaLeuGlnAspGlyLysLeuValTyr 380
QY 1174 ACCGAGTCCAAATATGACGCTACTACGCTTTCAGCAACACAGCGGACTTGTCTGGCCC 1233
Db 381 ValSerCysSerThrAlaValTyrGluAlaMetGlnMetAsp-----TirpMet 396
QY 1234 AACTTCATCGAA---GACCTCGAGGAGACTCTGTCTCTCCCGTGGTGTCTCCCTCATC 1290
Db 397 ArgAsnLeuGluValGlyIleProAlaLeuLeuGlnAspGlyLysLeuValTyr 416
QY 1291 TATGGCGAGCCGCTATCTGTCAACTGTGTGGCGGTGAGCGGCTTCCCTCGCTGCG 1350
Db 417 AlaGlyGluTyrAspLeuIleCysAsnTirpLeuGlyAsnSerLysTirpValHisGluMet 436
QY 1351 AACTACTCCCAAGCCCGCCAGTTCGAGCGGAGGTATACAGCCCTGAAAGTCAACGGC 1410
Db 437 GluTirpSerGlyGlnLysGluPheValAlaAlaThrValProPheHisValGlyAsn 456
QY 1411 GTCGAGTATGGGAAACTCGGAGTATGTTATTTCTCTCACTCGCTCTATGAGGCA 1470
Db 457 LysGluAlaGlyLeuMetLysAsnTyrGlySerLeuThrPheLeuLysValHisAspAla 476

```



```

Db 1826 ProArgThrProTyrGlyGluThrTrpThrGlyIleAsnTyrGluSerSerAspPheLeu 1845
QY 1123 -----TTTCTGCAAGGACTCTGTC 1143
Db 1846 AsnGlyTyrProCysTyrAspAspSerAlaMetGluGlyTyrLeuAsnArgProValVal 1865
QY 1144 ATGGACGCTATCGGCTC-----AACATCAACTACACCGCTCAATAATGACGTCTAC 1197
Db 1866 ArgLysAlaLeuAsnIleProAspSerValProTyrTyrAlaAlaAsnAsnIleIle 1885
QY 1198 TAGCGTTTCCAGCAACAGGCGCTTGTCTGCGCCCACTTCATCGAAGACCTCGAGGAG 1257
Db 1886 AsnAlaTyrAsnGlnValAspSerIleValProAsn-----LeuGlnIle 1901
QY 1258 ATCCCTGCT-----CTCCCCGTGGTGTCTCCCTCATCTATGGCGAGCGCATATAC 1308
Db 1902 IleMetAlaAsnAlaProAlaAsnPheLysMetLeuLeuTyrSerGlyAspAlaAspThr 1921
QY 1309 ATCTGCAACTGGTTCGGCGGTACG-----GCCGTTTCCCTC 1344
Db 1922 MetValAsnTrpLeuGlyAlaGluIlePheThrAlaAsnAsnPheAlaAlaLeuGlyLeu 1941
QY 1345 GCTCGAAGTACTCCAGCGGCC-----CAGTTCGGAAGCGGAGGTACAGGCC 1395
Db 1942 ThrThrSerProArgAlaGlnTrpThrPheGlnTyrAsnSerThr---PheGlnPro 1960
QY 1396 CTGAAGTCAACGGCGTCTGAGTATGGGAAACTCGCGAGTATGGTAATTCTCTCTCACT 1455
Db 1961 ---ThrValAlaGlyTyrGlnThrSerTyrThrSerAsnAlaIleAsnIleAspValLeu 1979
QY 1456 CGCTCTATGAGGCGCATGAGTCCATACCTACCGACCCATCGCTCCCTGCAATG 1515
Db 1980 ThrValLysGlySerGlyHisPheValProLeuAspArgProGlnGlnAlaLeuGlnMet 1999
QY 1516 TTAAACGGACTACTTCGGTGGATATCCAGAGGGCCAGAGAGATCTGCCCGCAGC 1575
Db 2000 -----IleTyrAsnPheValLysSerArgGlyTyrAsnThrProPheAspLeu 2015
QY 1576 TACAAGCAATGAACGGCTACAGCTACGATACAGATACAGTCTCGTGGCGCG-----CTG 1629
Db 2016 AsnSerAsnPheThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1630
QY 1630 CCTACGGCTACGACG 1644
Db 2036 ProThrValThrAla 2040

RESULT 14
Q9FMX9 PRELIMINARY; PRT; 499 AA.
AC Q9FMX9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Serine carboxypeptidase II-like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RL DNA Res. 4:401-414(1997).
DR EMBL; AB007648; BAB11176.1; -.
DR HSSP; P08819; 1WHT.
DR MEROPS; S10.005; -.
DR InterPro; IPR001563; Serine_carbpept.

```

```

DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine carbpept; 2.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
KW Carboxypeptidase
SQ SEQUENCE 499 AA; 56329 MW; 6150FB81A2E7BCBD CRC64;

```

Alignment Scores:

```

Pred. No.: 6,31e-21 Length: 499
Score: 411.50 Matches: 152
Percent Similarity: 41.06% Conservative: 80
Best Local Similarity: 26.90% Mismatches: 216
Query Match: 13.59% Indels: 117
DB: 10 Gaps: 26

```

US-09-712-338-1 (1-1668) x Q9FMX9 (1-499)

```

QY 13 GAATTTCTCTCAGTGTACCCCTTGGTTGCACGCCAGTTGGGCCCTTCCAGGAAGTACACG 72
Db 7 GluPhe---SerValLeuValLeuPheLeuValSerPheLeuLeu---GlySerThrSer 24
QY 73 GCGTCCGTGCGTAGAGACAGCTACCCAGAACCCACCACCGG-----GTCAAG 120
Db 25 Ala-----GluLysLeuCysSerAspAsnAspGlyAspAsnGlyCysPheArg 40
QY 121 ACTCTTCAACCGCAAAACAATGTCCACATCCGGTACAAGAA---CCCGGGCAGAGGCG 177
Db 41 SerArgValLeuAlaAlaGlnArgAlaAspArgValLysGluLeuProGly----- 57
QY 178 GTCTGCGAGACTACCCGGGTGTCAA-----TCCTACTCTGGATATGTGACACCTCT 231
Db 58 -----GlnProProValLysPheArgGlnTyrAlaGlyTyrVal---ThrVal 72
QY 232 CCCGAGTCCCAT-----ACCTTCTCTGGTTCCTCGAAGCAGACATACCCAGAA 282
Db 73 AsnGluThrHisGlyArgAlaLeuPheTyrTrpPheGluAlaThrGlnAsnProSer 92
QY 283 ACTGCACCTATCATTTGTGTGAATGGTGGCCCTCGAAGCGATTCTTTG---ATCGGT 339
Db 93 LysLysProValLeuLeuTyrPheAsnGlyGlyProGlyCysSerSerIleGlyPheGly 112
QY 340 CTCTTCAAGAGTTGGCCCTTGCCTATGTCATTCGACTTTTGTATGAC-----TACATC 393
Db 113 AlaAlaGluLeuGlyProPhePheProGlnAsnSerSerGlnProLysLeuLysLeu 132
QY 394 AACCTCCTCCTCGTGAACGAGGTCTCAANTTTACTATTCCTGCTCCGACCATGGGAGTC 453
Db 133 AsnProTyrSerTrpAsnLysAlaAlaAsnLeuPheLeuGluSerProValGlyVal 152
QY 454 GGCTTTTCATATAGTATGATACGTTGTATGGTTCCTTAAACCTGTAACTGGGGTCGTGAA 513
Db 153 GlyPheSerTyrThrAsnThr----- 159
QY 514 AATTCGAGCTTTCAGAGTTTCAGGCGCGGTACCCCAACCATTCATGCCATCTGATCGAT 573
Db 160 -----SerArgAspIleLysGlnLeuGly 167
QY 574 ACTACCAATCTTCCGAGAGCGGCTTGGGAGATCTCGAAGGATTCCTTAGTGACTA 633
Db 168 AspThr---ValThrAlaArgAspSerTyrAsnPheLeuValAsnTrpPheLysArgPhe 186
QY 634 CTTAGCTTGGACTCTAGGCTGAGTCTAAGGACTTCAGTCTATGGACGAGGAGCTATGGA 693
Db 187 Pro-----GlnTyrLysSerHisAspPheTyrIleAlaGlyGluSerTyrAla 202
QY 694 GGGCACTATGCTTCATCTTCATCATCTTTTCGAGCAGAATGAGAGAATTGCCAAC 753
Db 203 GlyHisTyrValProGlnLeuSerGluLeuIleTyrLysGluAsnLysIleAlaSerLys 222
QY 754 GGTAGTGTTAATGGTCTCAGCTTAATTCACATCTCTCGGAATTAATTACCGCATCATC 813

```

Db 223 LysAspPhe-----IleAsnLeuLysGlyLeuMetIleGlyAsnAlaLeuLeu 238
 QY 814 GACGAGCGATCCAGCGCCCTTACTACCTGCTGAATTCGTGTGACCAAAATACCTACGGTATC 873
 Db 239 AspAspGluThrAspGlnLysGlyMetIleGluTyrAlaTyrAspHisAlaValIleSer 258
 QY 874 AAGGCTGTCAAGAGACCGCTTACAACTACATGAAGTTTGGCAACCAAAATG---CCAAAT 930
 Db 259 AspAlaLeuTyrGluLysValAsnLysAsnGlyHisGlnLysLeuValThrLys 278
 QY 931 GGTTCGACGATTTGATTTCC----- 951
 Db 279 GluCysAsnAspAlaLeuAspGluTyrPheAspValTyrLysIleLeuAspMetTyrSer 298
 QY 952 -----ACCTGCAACAGACAAACCGCACCGCATTAGCTACCTACCTACCTTCGTCG 999
 Db 299 LeuTyrAlaProLysCysValProThrSerThrAsnSerSerThrSerHisSerValAla 318
 QY 1000 GCCAAGCCACCAACATGTGCGAGGACAATGTTGAGGGGCACTACTACGCGCTTTGCTGGT 1059
 Db 319 GlyAsnArgProLeuProAlaPheArgSerIleLeuArgProArgLeuIleSerHisAsn 338
 QY 1060 CGTGTGTGTATGATTCGGCATCCATGATGACCGGACTCCGCCAAGTTATTACAAC 1119
 Db 339 GluGlyTyrArgArgMetAlaGlyTyr---AspProCysAlaSerGluTyrThrGlu 357
 QY 1120 AAATTTCTGCAAGGACTCTGTGACGAGCTGTGCGGCTCAACATCACTCACTACACCCAG 1179
 Db 358 LysTyrMetAsnArgLysAspValGlnGluAlaLeuHisAlaVal----- 373
 QY 1180 TCCAAATAGCTCTACTAGCTTTCCAGCAACAGCGGACTTTGTC-----TGG--- 1230
 Db 374 -----ThrAsnIleSerTyrProTyrThrHisCysSerAspThrValSerPheTyrSer 391
 QY 1231 -----CCCACTTCAGAGAGACTCGAGAGATCTGTCTCCCGCGCGTGC 1281
 Db 392 AspAlaProLysMetLeuProThrLeuArgThrLeuValSerAlaGlyLeuArgVal 411
 QY 1282 TCCCTCATCTATGCGGCGGATTCATCATCTGCAACTGTTGCGGCGTCAAGCGGTTTCC 1341
 Db 412 TrpValPheSerGlyAspThrAsp-----GlyArgIleProVal 424
 QY 1342 CTGCTCGCAACTACTCC-----CAAGCGCCGAGTTCGAGCGCA 1393
 Db 425 ThrAlaThrArgTyrSerLeuLysLysLeuGlyLeuLysIleValGlnAspTyrThrPro 444
 QY 1384 GGGTACAGCCCTGAAAGTCAACGGGCTCGAGTATGGGAAACTCGCGAGTATGGTAAT 1443
 Db 445 TrpTyrThrLysLeuGlnValGly-----GlyTyrThrValGluTyrAspGly 460
 QY 1444 TTCTCTTCACTCGCGTCTATGAGCGAGCGCATGAGTCCCATACTACCGCCCATCGCC 1503
 Db 461 LeuMetPheValThrValArgGlyAlaGlyHisGlnValProThrPheLysProArgGlu 480
 QY 1504 TCCCTGCAATTTTACCGGACTATCTTCGTTGGGATATCGCAGAGGGCCAGAGAG 1563
 Db 481 AlaLeuGlnLeuIleHisPheLeu-----GlyAsnLysLys 493
 QY 1564 ATCTGGCCGAGTAC 1578
 Db 494 Leu---ProThrPhe 497
 RESULT 15
 Q9FYP7
 ID Q9FYP7 PRELIMINARY; PRT; 490 AA.
 AC Q9FYP7
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Similar to Hordeum vulgare carboxypeptidase D (putative
 DE carboxypeptidase D).
 GN P0011G08.39.
 OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0433F09.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0011G08.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF002539; BAB08188.1; -;
 DR EMBL; AF003225; BAB64666.1; -;
 DR HSSP; P08819; 1WHT.
 DR MEROPS; S10.005; -;
 DR Gramene; Q9FYP7; -;
 DR InterPro; IPR001563; Serine_carbpept.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF00450; serine_carbpept_1.
 DR PRINTS; PR00724; CRBOXIPTASEC.
 DR ProDom; PD001189; Serine_carbpept_1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 KW Carboxypeptidase.
 SQ SEQUENCE 490 AA; 54733 MW; 968FD450523D3574 CRC64;
 Alignment Scores:
 Pred. No.: 4.88e-19 Length: 490
 Score: 385.00 Matches: 125
 Percent Similarity: 41.77% Conservative: 68
 Best Local Similarity: 27.06% Mismatches: 179
 Query Match: 12.72% Indels: 90
 DB: 10 Gaps: 19
 US-09-712-338-1 (1-1668) x Q9FYP7 (1-490)
 QY 208 TACTCTGATAT-----GTGACACCTCTCCGAGTCCCATCTTCTTCTGTTCTTC 261
 Db 56 TyrSerGlyTyrValThrValAspAlaAlaGlyArgAlaLeuPheTyrTrpLeuIle 75
 QY 262 GAAGCCAGACATAACCCAGAAACTGCACCTATCACATTGTGGTTGAATGTGGCCCTGGA 321
 Db 76 GluAlaAla---AspProAlaSerAlaProLeuValLeuTrpLeuAsnGlyGlyProGly 94
 QY 322 AGCATCTTTTG---ATCGGTCTCTTCGAGAGTGGGCGCTGCCATGTCAATTCGACT 378
 Db 95 CysSerSerValGlyTyrGlyAlaSerGluGluLeuGlyAlaPheArgIleAsnProAsp 114
 QY 379 TTTCATGAC---TACATCAACCTCTCACTCGTGGAGAGGCTCCCAATTTACTATTCTG 435
 Db 115 GlyArgSerLeuTyrLeuAsnProTyrProTyrPheValAlaAlaAsnMetLeuPheLeu 134
 QY 436 TCCCGAGCATTTGGAGTTCGGCTTTTCATATAGTATGATACGTTGTGATGGGTTCATTAACCT 495
 Db 135 AspSerProAlaGlyValGlyTyrSerTyrSerAsnThrThrSerAspLeuPheThr 153
 QY 496 GTAACTGGGTCGTCGAAATTCGAGCTTTCGAGAGTTCAGGGCGGTCGCCAACCAT 555
 Db 154 -----AlaGly----- 155
 QY 556 GATGCCACTCTGATGATACCAATCTTCCCGCAGAGCGGCTTGGAGATCTCGCAA 615
 Db 156 -----AspAsnLysThrAlaHisAspSerTyrAlaPheLeuVal 168
 QY 616 GGATTCCTTAGTGGACTACCTAGCTTGGAGTCTTAGGGTGAGTCTTAAGAGCTTCACTA 675
 DB: 19


```
Db 169 AsnTrpLeuGluArgPhePro-----GlnTyrLysTyrArgAspPheTyrIle 184
QY 676 TGGACGGAGAGCTATGGAGGCACCTATGCTCTCCTCAATCTTTTACAGCAG 735
Db 185 AlaglyGluSerTyrGlyHisTyrValProGlnLeuSerGlnLeuValTyrArgAsn 204
QY 736 AATGAGAGAAATTGCCAAGGTAGTGTAAANGTGTTCAGCTTAATTTCAACTCTCTGGGA 795
Db 205 AsnLys-----AspValGluLysProIleLeuAsnPheLysGlyPheMet 219
QY 796 ATTATTAAACGCATCATCGACGGCGATCCAGGCCCTTACTACCTCGAATTCGCTGTG 855
Db 220 ValGlyAsnAlaValIleAspAsp-----TyrHisAspTyrValGlyThrPheGluTyr 237
QY 856 AACAAATACCTACGGTATCAAGGCTGTCAACGAGACCGTCTACAACTACATCAAGTTTGGC 915
Db 238 TrpTrpThrHisGlyLeu-----IleSerAspThrTyrGlnLysLeuGlnValAla 255
QY 916 AACCAATGCCAAT-----GGTTGCCAGGATTG-----ATTTC 951
Db 256 CysAspPheGluSerSerAlaHisAlaSerGluAlaCysAsnLysIleTyrGluValAla 275
QY 952 ACCTGGAACACAGACACACCGCATAGCTAGCTACGCCCTCTCGCGCGAAGCCACC 1011
Db 276 GluAlaGluGlnGlyAsnIleAspAlaTyrSerIleTyrThrProThrCysLysLysThr 295
QY 1012 AACATGTGCAGGACAAATGTTGAGGGCCATCTACGCCCTTGTGCTGTGTGTAT 1071
Db 296 SerPheLeuLysArgArgLeuIleArgGlyAsnSerProTrpLeuProArgGlyTyr--- 314
QY 1072 GATATCGGCATCCATATGATGACCGGACTCCGCCAAGTTATTACAAATAATTTCTGGCA 1131
Db 315 -----AspProCysThrGluLysTyrSerThrLysTyrTyrAsn 327
QY 1132 AAGGACTCTGTATGACGCTATCGGCGTCAACATCAAC-----TACACC 1176
Db 328 LeuProGluValGlnLysAlaPheHisAlaAsnValThrGlyIleProTyrAlaTrpThr 347
QY 1177 CAGTCCAATAATGACGTCTACTACGCTTTCAGCAACACAGCGACTTGTCTGGCCCAAC 1236
Db 348 ThrCysSerAspAspLeuPheTyrTyrTrpLysAspSerProArgSerMetLeuProIle 367
QY 1237 TTCAATCGAAGACCTCGAGGAGATCCTGTCTCTCCCGTGTCTCCCTCATCTATGGC 1296
Db 368 Tyr-----ArgGluLeuIleAlaAlaGlyLeuArgIleTrpValPheSerGly 383
QY 1297 GAGCGGATTACATCTGCAACTGTTCGGCGGTGAGCGGCTTCCCTC---GCTGCGAAC 1353
Db 384 AspAlaAspSerVal-----ValProLeuThrAlaThrArg 395
QY 1354 TACTCCCAAGCGCCAGTTCCGAAGCGGAGGTACACGCCCTGAAAGTCAACGCGGTC 1413
Db 396 TyrSer-----IleAspAlaLeuTyrLeuProThrValThrAsnTrpTyr 410
QY 1414 GAGTATGGGAAACTCGCAG-----TATGGTAATTTCTCTTC 1452
Db 411 ProTrpTyrAspAspGluGluValAlaGlyTyrCysGlnValTyrGlnGlyLeuThrLeu 430
QY 1453 ACTCGGCTATGAGCGACCATGAGTCCCATCTACAGCCCATCGCTCCCTGCAA 1512
Db 431 ValThrIleArgGlyAlaGlyHisGluValProLeuHisArgProArgGlnAlaLeuLys 450
QY 1513 TTGTTT 1518
Db 451 LeuPhe 452
```

Search completed: November 21, 2003, 17:44:19
Job time : 114.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2003, 20:31:12 ; Search time 579 Seconds

(without alignments)
9415.138 Million cell updates/sec

Title: US-09-712-338-1

Perfect score: 1668

Sequence: 1 atgctgggtacgaattctt.....ccagtggtgtatgcatag 1668

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09D_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.6	4.2	2218	14	US-10-206-619-1
2	56.4	3.4	1814	10	US-09-736-457-319
3	56.4	3.4	1814	10	US-09-902-941-319
4	56.4	3.4	1814	10	US-09-849-626-319
5	56.4	3.4	1814	11	US-09-476-300-319
6	56.4	3.4	1814	12	US-10-113-872-319
7	56.4	3.4	1814	14	US-10-017-754-319
8	56.4	3.4	1815	10	US-09-880-107-2296
9	54.8	3.3	2864	9	US-09-925-302-331
10	47.6	2.9	390	9	US-09-790-399-7
11	47.6	2.9	1551	13	US-10-084-018-4
12	47.6	2.9	1684	9	US-09-729-674-109
13	47.6	2.9	1697	11	US-09-796-753-39
14	47.6	2.9	1949	12	US-09-971-429B-33
15	47.6	2.9	2076	10	US-09-909-320-163
16	47.6	2.9	2076	10	US-09-909-088B-163

17	47.6	2.9	2076	10	US-09-905-291A-163	Sequence 163, App
18	47.6	2.9	2076	10	US-09-902-853-163	Sequence 163, App
19	47.6	2.9	2076	10	US-09-907-824-163	Sequence 163, App
20	47.6	2.9	2076	10	US-09-907-841-163	Sequence 163, App
21	47.6	2.9	2076	11	US-09-904-011-163	Sequence 163, App
22	47.6	2.9	2076	11	US-09-906-742-163	Sequence 163, App
23	47.6	2.9	2076	11	US-09-906-838-163	Sequence 163, App
24	47.6	2.9	2076	11	US-09-907-613-163	Sequence 163, App
25	47.6	2.9	2076	11	US-09-907-942-163	Sequence 163, App
26	47.6	2.9	2076	11	US-09-904-859-163	Sequence 163, App
27	47.6	2.9	2076	11	US-09-909-204-163	Sequence 163, App
28	47.6	2.9	2076	11	US-09-904-820-163	Sequence 163, App
29	47.6	2.9	2076	11	US-09-904-786-163	Sequence 163, App
30	47.6	2.9	2076	11	US-09-906-646-163	Sequence 163, App
31	47.6	2.9	2076	11	US-09-906-700-163	Sequence 163, App
32	47.6	2.9	2076	11	US-09-903-786-163	Sequence 163, App
33	47.6	2.9	2076	11	US-09-902-903-163	Sequence 163, App
34	47.6	2.9	2076	11	US-09-903-749A-163	Sequence 163, App
35	47.6	2.9	2076	11	US-09-904-119-163	Sequence 163, App
36	47.6	2.9	2076	11	US-09-904-956-163	Sequence 163, App
37	47.6	2.9	2076	11	US-09-902-736-163	Sequence 163, App
38	47.6	2.9	2076	11	US-09-907-794-163	Sequence 163, App
39	47.6	2.9	2076	11	US-09-903-943-163	Sequence 163, App
40	47.6	2.9	2076	11	US-09-904-462-163	Sequence 163, App
41	47.6	2.9	2076	11	US-09-907-925-163	Sequence 163, App
42	47.6	2.9	2076	11	US-09-902-692-163	Sequence 163, App
43	47.6	2.9	2076	11	US-09-903-520-163	Sequence 163, App
44	47.6	2.9	2076	11	US-09-905-056-163	Sequence 163, App
45	47.6	2.9	2076	11	US-09-909-064-163	Sequence 163, App

ALIGNMENTS

RESULT 1

US-10-206-619-1

; Sequence 1, Application US/10206619

; Publication No. US20030108908A1

; GENERAL INFORMATION:

; APPLICANT: Korea Research Institute of Bioscience and Biotechnology

; APPLICANT: Dong Kook Pharmaceutical Co.

; APPLICANT: Rhee, Sangki

; APPLICANT: Choi, Eulsung

; APPLICANT: Kang, Hyunah

; APPLICANT: Sohn, Junghoon

; APPLICANT: Bae, Junghoon

; APPLICANT: Kim, Moowoong

; APPLICANT: Agaphonov, Michasel

; TITLE OF INVENTION: Hansenula polymorpha mutants and process for the preparation of recombinant proteins using the same

; FILE REFERENCE: 4220-116 US

; CURRENT APPLICATION NUMBER: US/10/206,619

; CURRENT FILING DATE: 2002-07-26

; PRIOR APPLICATION NUMBER: US/09/674,617

; PRIOR FILING DATE: 2001-01-03

; PRIOR APPLICATION NUMBER: PCT/KR00/00173

; PRIOR FILING DATE: 2000-03-04

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 2218

; TYPE: DNA

; ORGANISM: Hansenula polymorpha D11

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: U67174

; DATABASE ENTRY DATE: 1996-08-17

; RELEVANT RESIDUES: (1)..(2218)

US-10-206-619-1

Query Match 4.2%; Score 70.6; DB 14; Length 2218;
Best Local Similarity 53.4%; Pred. No. 4.3e-12;
Matches 148; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

241 CATACCTTCTCTGTTCTTTCAGCCAGACATACCCAGAAACTGCACCTATCATATG 300
|||
840 CATTTCTTCTATTTGATGTTTGGTCCAGAAATGACCGGTCAACGACCTGTGATTCG 899
|||
301 TGGTTGAATGGTGGCCCTGGAGCGAATCTTTTGATCGTCTCTTTCAGAGATGGGCCCT 360
|||
900 TGGCTCAACGGGTGGTCCAGGATGCTCTTCTTGAAGTGAATGCTTTTGGAGCTGGCTCT 959
|||
361 TGCCATGTCATTCGACTTTTTCATGACTACATCAACCTCTCACTCGTGGAAACGAGGTCTCC 420
|||
960 GCTTCTATCGTCCAGATCTCAAGCAATCAACACCATATTCGTGGAAATCCATGCC 1019
|||
421 AATTTACTATTCCTGTCCAGCCATTTGGAGTCGGCTTTTCATATAGTATACGGTTGAT 480
|||
1020 ACTGTGATTTTCTTTCAGCCAGCTGTCAATGTTGATTCGTCTCTCTTCCAGTCTGTT 1079
|||
481 GGGTCATTAACCTGTACTGGGTCTGCGAAATTT 517
|||
1080 TCTAACCGTCCGAGCTGGTAAAGACGCTATGCTT 1116
|||

RESULT 2

US-09-736-457-319

; Sequence 319, Application US/09736457

; Patent No. US20020168637A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Tonglong

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vedwick, Tom

; APPLICANT: Carter, Darrick

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; APPLICANT: Fan, Liqun

; APPLICANT: Wang, Aljun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.478C15

; CURRENT APPLICATION NUMBER: US/09/736,457

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 1864

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 319

; LENGTH: 1814

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-736-457-319

Query Match 3.4%; Score 56.4; DB 10; Length 1814;
Best Local Similarity 52.5%; Pred. No. 2.2e-07;
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

QY 193 CCGGGTGTCAATCTTCTGATATGTCGACACTCTCCCGAGTCCCATCTCTTC 252
|||
Db 139 CGCTCTTCCGCGAGTACTCCGCTACCTCAAAAGCTCCGGCTCCAGCACCTCCACTAC 198
|||
QY 253 TGGTTCTTTCGAAGCCAGACATAACCCAGAAACTGCACCTATCACATTTGGTTGAATGGT 312
|||
Db 199 TGGTTTGTGGAGTCCCGAGAGGATCCGAGAACAGCCCTGTGGTCTTGGCTCAATGGG 258
|||
QY 313 GGCCCTGGAGCGAATCTTTGATCGGTCTCTTTCAGAGATGGGCCCTTGCATATGCAAT 372
|||
Db 259 GGTCCCGGTGTCAGCTCACTAGATGGGCTCCTCACAGCATGGCCCTTCTCTGGTCCAG 318
|||
QY 373 TCGACTTTTGTATGACT---ACATCAACCTCTCACTCGTGGAAACGAGGTCTCCAAATTTACTA 429
|||
Db 319 CCAGATGGTGTACCCCTGGAGTACAAACCTATCTTGGAACTGTGATTCGAATGTGTTA 378
|||
QY 430 TTCTCTCCAGCCATTTGGAGTCCGGCTTTTCATATAGTATGAT 471
|||
Db 379 TACCTGGAGTCCCGAGCTGGGTGGGTCTTCTCTACTCCGAT 420
|||

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 319
; LENGTH: 1814
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-849-626-319

```

Query Match	3.4%;	Score 56.4;	DB 10;	Length 1814;
Best Local Similarity	52.5%;	Pred. No. 2.2e-07;		
Matches 148;	Conservative	0;	Mismatches 131;	Indels 3; Gaps 1;

QY	193	CGGGGTCTCAAACTCTACTCTGGATATGTGGACACCTCTCCGAGTCCCATCTTCTC	252
Db	139		
		CCGTCTTTCGGCAGTACTCCGGTACTCTCAAAAGCTCCGGCTCCAAGCACCTCCACTAC	198
QY	253	TGTTTCTTCAAGCSGACACATAACCCAGAAACTGCACCTATCATCTTGTTGGTTCGAATGGT	312
Db	199		
		TGTTTTGTGGAGTCCCAAGAGGATCCCGAACAACGCCCTGTGSGCTTTGGCTCAATGGG	258
QY	313	GGCCCTCGGAAGCGATTCTTTGATCGGTCTCTTCGAAGAGTTGGSCCTTCGCATGTCAT	372
Db	259		
		GTCTCCGGCTGCAGCTCACTAGATGGCTCTCTACAGAGCATGCCCCCTTCTTGGTCCAG	318
QY	373	TCGACTTTTGAAGACT---ACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTACTA	429
Db	319		
		CCAGATGGTGTCAACCCTGGAGTCAACCCCTATCTTGGHAATCTGATTGCCAATGCTTA	378
QY	430	TTCTCTGCCAGCATGGGAGTCGGCTTTTCATATAGTCAT	471
Db	379		
		TACCTGGAGTCCCAAGCTGGGGTGTCTCTACTCCGAT	420

RESULT 5

US-09-476-300-319
; Sequence 319, Application US/09476300
; Publication No. US20030125245A1
; GENERAL INFORMATION:

APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER

```

; FILE REFERENCE: 210121.478C3
; CURRENT APPLICATION NUMBER: US/09/476,300
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 785
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 319
; LENGTH: 1814

```

LENGTH: 1614
TYPE: DNA
ORGANISM: Homo sapiens
US-09-476-300-319

Query Match	3.4%	Score 56.4;	DB 11;	Length 1814;
Best Local Similarity	52.5%;	Pred. No. 2.2e-07;		
Matches 148;	Conservative 0;	Mismatches 131;	Indels 3;	Gaps 17;

193	CGGGTGTCAAATCCTACTCTGTGATATGTGGACACCTCTCCGAGTCCCATACCTTCTTC	252
139	CCGTCTTTCGCCAGTACTCCGGCTACCTCAAAGCTCCGGTCCAAGACACTCACTAC	198
253	TGGTCTTCGAAGCCACACATACCCAGAACTGCACCTATCACATGTGGTGTGAATGGT	312
199	TGGTTTGTGGAGTCCCAAGAGGATCCGGAACAGGCCCTGTGCTGTTTGGCTCAATGGG	258
313	GGCCTGGAAGCGGATCTTTGATCGGTCTCTTCCAAAGAGTTGGGCCCTTCCCATGTCAT	372
259	GGTCCGGCTCAGCTCACTAGATGGGCTCTCACAGAGCATGGCCCCCTCTCTGTTCAG	318
373	TCGACTTTTGATGACT---ACATCAACCTCACTCGTGGAAAGAGTGTCCAAATTTACTA	429
319	CCAGATGGTGTCACTCCCTGGAGTACAAACCCCTATTTCTTGGAACTGATTCGCAATGTGTA	378
430	TTCTGTGCCAGCAATGGGAGTCGGCTTTTCATATAGTAT	471

Db 379 TACCTGGAGTCCCAGCTGGGGTGGGCTTCTCCTACTCCGAT 420

RESULT 6

```

US-10-113-872-319
; Sequence 319, Application US/10113872
; Publication No. US20030170255A1
; GENERAL INFORMATION:
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C19
; CURRENT APPLICATION NUMBER: US/10/113,872
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: PascSeq for Windows Version 4.0
; SEQ ID NO 319
; LENGTH: 1814
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-113-872-319

```

Query Match	3.4%	Score 56.4	DB 12	Length 1814
...

Best Local Similarity	52.5%;	Pred No. 2.2e-07;	Length 104;
Matches	148;	Conservative	0;
		Mismatches	131;
		Indels	3;
		Gaps	1;
QY	193	CGGGTGTCAAATCTACTCTGGATATGTCGACACCTCTCCGAGTCCCATACCTTCTTC	252
Db	139	CGCTCTTCCGCGAGTACTCCGGTACCTCAAAAGCTCCGGCTCCAGCACCTCCACTAC	198
QY	253	TGGTCTTTCGAAGCCAGACATATACCAGAAATGCACTATCATTTGTTGTTGAATGGT	312
Db	199	TGGTTTGTGGAGTCCCGAAGAGATCCCGAAACAGCCCTGTGGTGCTTTGGTCAATGGG	258
QY	313	GGCCCTCGAAGCGATCTTTGATCGTCTCTTCGAAGAGTTGGCCCTTGCCTATGTCAAT	372
Db	259	GGTCCGGCTGACGTCACTAGATGGGCTCCTCACAGAGCATGCCCTCTCTCGTGCCAG	318
QY	373	TCGACTTTTGTATGACT---ACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTACTA	429
Db	319	CCAGATGGTGTCAACCCTGGAGTACCAACCCCTATCTTGGNATCTGATGTCGAATGTGTTA	378
QY	430	TTTCCGTCCAGCCATTGGAGTCGGCTTTTCATATAGTGAAT	471
Db	379	TACCTGGAGTCCCAAGCTGGGGTGGGCTTCTCCTACTCCGAT	420

RESULT 7

US-10-017-754-319
; Sequence 319, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C18

313 GGCCCTGGAAGCGATTCTTTGATCGGTCTCTTTCGAAGAGTTGGCCCTTCCCATGTCAAT 372
259 GGTCCCGGCTGCAGCTCACTAGATGGGCTCTCAGAGAGCATGGCCCTTCTCTGGTCCAG 318
373 TCGACTTTTGTATGACT---ACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTACTA 429
319 CCAGATGGTGTACCCCTGGAGTACACCCCTATCTTGGATCTGATGTCGAATGTGTTA 378
430 TTCTGTCTCCAGCCATTGGGAGTGGCTTTTTCATATAGTGAT 471
379 TACCTGGAGTCCCGAGCTGGGTGGGCTTCTCTACTCCGAT 420

RESULT 9
US-09-925-302-331
; Sequence 331, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 331
; LENGTH: 2864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2850)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2858)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2860)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-331

Query Match 3.3%; Score 54.8; DB 9; Length 2864;
Best Local Similarity 52.1%; Pred. No. 9.7e-07;
Matches 147; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

193 CCGGGTGTCAATCCTACTCTGATATGTGGACACTCTCCGAGTCCCATACCTTCTTC 252
171 CCGTCTTTCGCGCAGTACTCCGGTCACTCAAGGCTCCGCTCCAGCACCTCCACTAC 230
253 TGGTTCTTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTTGGTTGAATGGT 312
231 TGGTTTGGAGTCCAGAGGATCCGAGACACCCCTGTGGTGTGGCTCAATGGG 290
313 GGCCCTGGAAGCGATTCTTTGATCGGTCTCTTTCGAAGAGTTGGGCCCTTCCCATGTCAAT 372
291 GGTCCCGGCTGCAGCTCACTAGATGGGCTCTCAGAGAGCATGGCCCTTCTCTGGTCCAG 350
373 TCGACTTTTGTATGACT---ACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTACTA 429
351 CCAGATGGTGTACCCCTGGAGTACACCCCTATCTTGGATCTGATGCCAATGTGTTA 410
430 TTCTGTCTCCAGCCATTGGGAGTGGCTTTTTCATATAGTGAT 471
411 TACCTGGAGTCCCGAGCTGGGTGGGCTTCTCTACTCCGAT 452

RESULT 10
US-09-790-399-7

313 GGCCCTGGAAGCGATTCTTTGATCGGTCTCTTTCGAAGAGTTGGCCCTTCCCATGTCAAT 372
259 GGTCCCGGCTGCAGCTCACTAGATGGGCTCTCAGAGAGCATGGCCCTTCTCTGGTCCAG 318
373 TCGACTTTTGTATGACT---ACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTACTA 429
319 CCAGATGGTGTACCCCTGGAGTACACCCCTATCTTGGATCTGATGTCGAATGTGTTA 378
430 TTCTGTCTCCAGCCATTGGGAGTGGCTTTTTCATATAGTGAT 471
379 TACCTGGAGTCCCGAGCTGGGTGGGCTTCTCTACTCCGAT 420

RESULT 8
US-09-880-107-2296
; Sequence 2296, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2296
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M22960
US-09-880-107-2296

Query Match 3.4%; Score 56.4; DB 10; Length 1815;
Best Local Similarity 52.5%; Pred. No. 2.2e-07;
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

193 CCGGGTGTCAATCCTACTCTGATATGTGCACTCTCCGAGTCCCATACCTTCTTC 252
139 CCGTCTTTCGCGCAGTACTCCGGTCACTCAAGAGTCCGCTCCAGCACCTCCACTAC 198
253 TGGTTCTTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTTGGTTGAATGGT 312
199 TGGTTTGGAGTCCAGAGGATCCCGAGAACAGCCCTGTGGTCTTTGGCTCAATGGG 258
313 GGCCCTGGAAGCGATTCTTTGATCGGTCTCTTTCGAAGAGTTGGGCCCTTCCCATGTCAAT 372
259 GGTCCCGGCTGCAGCTCACTAGATGGGCTCTCAGAGAGCATGGCCCTTCTCTGGTCCAG 318
373 TCGACTTTTGTATGACT---ACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTACTA 429
319 CCAGATGGTGTACCCCTGGAGTACACCCCTATCTTGGATCTGATGTCGAATGTGTTA 378
430 TTCTGTCTCCAGCCATTGGGAGTGGCTTTTTCATATAGTGAT 471
379 TACCTGGAGTCCCGAGCTGGGTGGGCTTCTCTACTCCGAT 420

RESULT 8
US-09-880-107-2296
; Sequence 2296, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2296
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M22960
US-09-880-107-2296

Query Match 3.4%; Score 56.4; DB 10; Length 1815;
Best Local Similarity 52.5%; Pred. No. 2.2e-07;
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

193 CCGGGTGTCAATCCTACTCTGATATGTGCACTCTCCGAGTCCCATACCTTCTTC 252
139 CCGTCTTTCGCGCAGTACTCCGGTCACTCAAGAGTCCGCTCCAGCACCTCCACTAC 198
253 TGGTTCTTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTTGGTTGAATGGT 312
199 TGGTTTGGAGTCCAGAGGATCCCGAGAACAGCCCTGTGGTCTTTGGCTCAATGGG 258

US-09-729-674-109
; Sequence 109, Application US/09729674
; Patent No. US20010039335A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Meirberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steinger II, Robert J.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6055-64X
; CURRENT APPLICATION NUMBER: US/09/729,674
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: 09/539,330
; PRIOR FILING DATE: 2000-03-30
; NUMBER OF SEQ ID NOS: 283
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 109
; LENGTH: 1684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-729-674-109

Query Match 2.9%; Score 47.6; DB 9; Length 1684;
Best Local Similarity 46.9%; Pred. No. 0.00018;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 220 GTCGACACCTCCCGAGTCCCATACCTTCTCTGGTCTTCGAGCCGAGCATACCCCA 279
DB 332 GTGATAAGACTTACACAGCAACCTCTCTCTGGTCTTCCAGCTCAGATACAGCCA 391
QY 280 GAACTGCACCTATCACATGTGGTGTGATGTGGCCCTGGAAGCGATCTTTGATCGGT 339
DB 332 GAAGATGCCCGAGTAGTTCTCTGGCTACAGGTGGCGGGAGGTTCATCCAGTTTGA 451
QY 340 CTCCTCGAAGAGTCTGGGCCCTTGGCATGTCAATTCGACTTTTGATCATCAACCCCT 399
DB 452 CTCCTTGGAGACATGGGCTTATGTTGTCACAGTAACATGACCTTGGTGCAGAGAC 511
QY 400 CACTCGTGAAGAGAGGTCTCCAAATTTACTATTCCTGCTCCAGCCATGGAGTGGCTTT 459
DB 512 TTCCCTGGACACACAGCGCTCTCCATGCTTTACATTTGACATCCAGTGGCGACAGCTTC 571
QY 460 TCATATAGTATACAGGTTCATGGTCCATTAACCTGTAACTGGGTGCTCGAAATTCG 519
DB 572 AGTTTACTGATATACCCAGCATATGAGTCAATGAGGACGATGATACCGGATTTA 631
QY 520 AGCTTTCAGGAGTTCAG 537
DB 632 TACAGTGCATTAATTCAG 649

RESULT 13
US-09-796-753-39
; Sequence 39, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30

; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 39
; LENGTH: 1697
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)...(1528)
US-09-796-753-39

Query Match 2.9%; Score 47.6; DB 11; Length 1697;
Best Local Similarity 46.9%; Pred. No. 0.00018;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 220 GTCGACACCTCCCGAGTCCCATACCTTCTCTGGTCTTCGAGCCGAGCATACCCCA 279
DB 338 GTGATAAGACTTACACAGCAACCTCTCTCTGGTCTTCCAGCTCAGATACAGCCA 397
QY 280 GAACTGCACCTATCACATGTGGTGTGATGTGGCCCTGGAAGCGATCTTTGATCGGT 339
DB 398 GAAGATGCCCGAGTAGTTCTCTGGCTACAGGTGGCGGGAGGTTCATCCATGTTGA 457
QY 340 CTCCTCGAAGAGTGTGGGCCCTTGGCATGTCAATTCGACTTTTGATCATCAACCCCT 399

Db 458 CTCTTTGTGAACATGGGCGCTTATGTTGTCAACAGTAACATGACCTTGGCTGACAGAGAC 517
QY 400 CACTCGTGGAAACAGAGTCTCCATTTACTATTCTGTCCAGCCATTGGAGTGGCGTTT 459
Db 518 TTCCCTGGACCAACAGCTCTCCATGCTTTACATTGACAAATCAGTGGGCAACAGGCTTC 577
QY 460 TCATATAGTATACGGTGTGATGGTCCATTAACCTGTAACTGGGTGCGTGAATAATCG 519
Db 578 AGTTTACTGATGATACCCACGATATGCAATGAGGACGATGATGACACGGGATTTA 637
QY 520 AGCTTTGACGAGTTCAG 537
Db 638 TACAGTGCCTAATTTCAG 655

RESULT 14
US-09-971-429B-33
; Sequence 33, Application US/09971429B
; Publication No. US20030175704A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K. W.
; APPLICANT: Shyjan, Christopher M.
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0040 US
; CURRENT APPLICATION NUMBER: US/09/971,429B
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/239,024
; PRIOR FILING DATE: 2000-04-10
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 33
; LENGTH: 1949
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175704A1 239680.4
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 71, 73
; OTHER INFORMATION: a, t, c, g, or other
US-09-971-429B-33

Query Match 2.9%; Score 47.6; DB 12; Length 1949;
Best Local Similarity 46.9%; Pred. No. 0.0002;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 220 GTCGACACCTCCGAGTCCCATACCTTCTTGTGTTCTTCGAGCCAGACATAACCCA 279
Db 608 GTGAATAAGACTTACACAGCAACCTCTTCTGTGTTCTTCCAGCTCAGATACAGCCA 667
QY 280 GAAACTGCACCTATCACATTGTGTTGAATGCTGCTGCTGAGGATCTTTGATCGGT 339
Db 668 GAAGATCCCGAGTAGTCTCTGCTACAGGTGGCGGGAGGTTTCATCATGTTTGA 727
QY 340 CTCTCGAAGAGTTGGGCGCTTCGATGTCATTAATTCGATTTTGATGACTATCAACCT 399
Db 728 CTCTTTGTGAACATGGGCTTATGTTGTCAAGTACATGACCTTGGCGTGACAGAGAC 787
QY 400 CACTCTGGAAAGAGGTCTCAATTTACTATTCCTGTCCAGCCATTCGGAGTCGGGTTT 459
Db 788 TTCCCTGGACCAACAGCTCTCCATGCTTTACATTGACAATCCAGTGGGCAACAGGTTTC 847
QY 460 TCATATAGTATACGGTGTGATGGTCCATTAACCTGTAACTGGGTGCGTGAATAATCG 519
Db 848 AGTTTACTGATGATACCCAGGATATGCAATGAGGACGATGATGACACGGGATTTA 907
QY 520 AGCTTTGACGAGTTCAG 537
Db 908 TACAGTGCCTAATTTCAG 925

RESULT 15
US-09-909-320-163
; Sequence 163, Application US/09909320
; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Iuc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076

Mon Nov 24 13:41:04 2003

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-909-320-163

Query Match      2.9%; Score 47.6; DB 10; Length 2076;
Best Local Similarity 46.9%; Pred. No. 0.00021;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 220 GTCGACACCTCTCCCGAGTCCCATACCTCTTCTGGTTCCTGGAAGCCAGACATAACCCA 279
Db 334 GTGAATAAGACTTACAAACAGAACCTCTTCTGGTTCCTGGAAGCCAGACATAACCCA 393
QY 280 GAAACTGCACCTATACATTTGATGGAATGGTGGCCCTGGAAGCCGATTCCTTTCATCGGT 339
Db 394 GAAGATGCCCCAGTAGTTCTCTGGCTACAGGTGGGCGGGAGGTTTCATCATGTTTGA 453
QY 340 CTCTTCGAAGAGTTGGCCCTTGCATGTCAATTCGACTTTTGATGACTACATCAACCCCT 399
Db 454 CTCCTTTGTGAACATGGGCCCTTATGTTGTCAAGTAACATGACCTTGGGTGACAGAGAC 513
QY 400 CACTCGTGGAAAGAGGTCTCCAAATTTACTATTCCTGCCAGCCATTGGGAGTCGGCTTT 459
Db 514 TTCCCTTGGACCACAAAGCTCTCCATGCTTTACATTCGATCCAGTGGGCACAGGCTTC 573
QY 460 TCATATAGTGTATACGGTTGATGGGTCCATTAACCCCTGTAACTGGGGTCGTGGAATAATCG 519
Db 574 AGTTTACTGATGATACCCACGGATATGCACTCAATGAGGACGATGTAGCACGGGATTTA 633
QY 520 AGCTTTGCAGGAGTTCAG 537
Db 634 TACAGTGCACATAATTCAG 651

```

Search completed: November 21, 2003, 14:38:11
Job time : 582 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 21, 2003, 17:27:39 ; Search time 40.5 Seconds
(without alignments)
7921.450 Million cell updates/sec

Title: US-09-712-338-1
Perfect score: 3027
Sequence: 1 atgcgtggtacgaatttct.....ccagtgtgtatggcatag 1668

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEU=frame_n2p.model -DEV=xlp
-Q/cgn2_1/USPTO spoal_p/US09712338/runat 17112003 170149 10540/app query.fasta_1.1863
-DB=PIR_76 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTWNT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09712338 @CGN 1 1 59 @runat 17112003 170149 10540 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_76.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	29.5	423	2	S38953
2	749	24.7	460	1	S51516
3	652.5	21.6	1002	2	T43236
4	636	21.0	508	1	S46008
5	601	19.9	552	2	JC7666
6	596	19.7	542	1	JC1380
7	594.5	19.6	523	1	S61713
8	527.5	17.4	532	1	CPBYV
9	473	15.6	429	1	T03607
10	455	15.0	500	1	S22530
11	439	14.5	510	2	T48977
12	425.5	14.1	411	1	A35275
13	417.5	13.8	2105	2	T18968
14	416.5	13.8	499	1	A29412

15	383	12.7	482	2	T49079
16	365	12.1	480	2	T50511
17	362.5	12.0	487	2	T49080
18	358	11.8	470	2	E96637
19	355.5	11.7	470	2	T16806
20	351	11.6	482	2	A43828
21	349	11.5	474	2	A35732
22	348	11.5	452	2	H84772
23	347.5	11.5	516	2	S44191
24	344	11.4	465	2	G86244
25	344	11.4	501	2	T49081
26	344	11.4	512	2	T33463
27	337.5	11.1	425	2	F85360
28	337.5	11.1	487	2	H84472
29	337	11.1	480	2	A31589
30	336	11.1	465	2	B85358
31	333.5	11.0	510	2	T39601
32	332	11.0	459	2	T16316
33	331	10.9	729	2	A29651
34	330.5	10.9	474	2	D84631
35	330	10.9	1203	2	T21275
36	328	10.8	2338	2	T25810
37	327.5	10.8	479	2	G96501
38	326.5	10.8	425	2	E84631
39	326.5	10.8	454	2	T20829
40	323	10.7	502	2	T49188
41	320.5	10.6	456	2	H86406
42	318	10.5	468	2	D84503
43	316	10.4	574	2	T16230
44	314.5	10.4	476	1	T05701
45	313	10.3	456	2	A85139

ALIGNMENTS

RESULT 1

S38953
carboxypeptidase D (EC 3.4.16.6) - Penicillium janthinellum
N/Alternate names: carboxypeptidase S1
C/Species: Penicillium janthinellum
C/Date: 18-Feb-1994 #sequence_revision 19-Apr-1996 #text_change 31-Mar-2000
C/Accession: S38953
R/Svendsen, I.; Hofmann, T.; Endrizzi, J.; Remington, S.J.; Breddam, K.
PES Lett. 333, 39-43, 1993
A/Title: The primary structure of carboxypeptidase S1 from Penicillium janthinellum.
A/Reference number: S38953; PMID:94039747; PMID:9224168
A/Accession: S38953
A/Status: preliminary
A/Molecule type: protein
A/Residues: 1-423 <SVB>
C/Superfamily: serine carboxypeptidase
C/Keywords: hydrolase; serine carboxypeptidase

Alignment Scores:

Pred. No.:	2.44e-51	Length:	423
Score:	894.00	Matches:	186
Percent Similarity:	54.31%	Conservative:	66
Best Local Similarity:	40.09%	Mismatches:	160
Query Match:	29.53%	Indels:	52
DB:	2	Gaps:	8

US-09-712-338-1 (1-1668) x S38953 (1-423)

QY	175	GGCGTCTGCAGACTACCCGGGTGCAATCTCTCGATATGTCGACACCTCTCCC	234
DB	6	GlyIleCysGluThrThrProGlyValAsnGlnTy-SerGlyTyLeuSerValGlySer	25
QY	235	GAGTCCCATACCTTCTTCTGTTCTTCGAGCCAGACATAACCCAGAACTGCACCTATC	294
DB	26	AsnMetAsnMetTrpPheTrpPheGluAlaArgAsnAsnProGlnAlaProLeu	45
QY	295	ACATTGTTGTTGAATGGTGGCCCTCGAAGCGATTTCTTTCATCGGTCTCTTCGAGAGTTG	354

Db 46 AlaAlaTIPheAsnGlyGlyProGlyCysSerMetIleGlyLeuPheGlnGluAsn 65
QY 355 GGCCTTGGCAT---GTCAATTCGACTTTTGATGACTACATCAACCTCACTCGTGGAAAC 411
Db 66 GlyProCysHisPheValAsnGlyAspSerThrProSerLeuAsnGluAsnSerTrpAsn 85
QY 412 GAGGTCTCAATTTACTATCTCTGCTCCAGCCATTGGAGTCGGCTTTTCATATAGTAT 471
Db 86 AsnTyAlaAsnMetIleIleAspGlnProIleGlyValGlyPheSerTyGly--- 104
QY 472 ACGGTTGATGGTCCATTAAACCTGTAACTGGGTGCTGCAAAATTCGAGCTTTGCAGGA 531
Db 104 ----- 104
QY 532 GTTCAGGGCGGTACCCACCATTTGATGCGCACTCTGATGATACCAATCTTGGCGCA 591
Db 105 -----ThrAspAspValThr-----SerThrValThrAlaAla 115
QY 592 GAGCGCGTTGGAGATCTCTGCAAGGATTCCTTAGTGACTTACCTAGCTGGACTTAGG 651
Db 116 ProTyValTrpAsnLeuGlnAlaPheTyAlaGlnArgProGluTyArgGluSerArg 135
QY 652 GTGCAGTCTAAGACTTCAGTCTATGACGGAGAGCTATGAGGGCACTATGCTCTGCA 711
Db 136 -----AspPheAlaIlePheThrGluSerTyGlyGlyHisTyGlyProGlu 151
QY 712 TTCTTCATCATTTTACGACAGATGAGAGATTCGCAACGGTAGTGTAAATGGTGT 771
Db 152 PheAlaSerTyIleGluGlnGlnAsnAlaIleAlaIleAlaGlySerValThrGlyGln 171
QY 772 CAGCTTAATTTCACTCTCTGGGAATTTAATACGGCATTCATCGAGCGGATCGAGGCC 831
Db 172 AsnValAlaIleValAlaGlyValAsnAsnGlyTrpIleAspSerThrIleGlnGlu 191
QY 832 CTTTACTACCTGAAATTCGCTGTGAACAATACCTACCGTATCAAGCTGTCAACGAGCC 891
Db 192 LysAlaTyIleAspPheSerTyAsnAsnSerTyGlnGlnIleAspSerSerThr 211
QY 892 GTCACACTACATGAGTTTCCCAACCAATGCGCAATGGTGTCCAGGATTTGATTC 951
Db 212 ArgAspSerLeuLeuAspAlaTyAsn-----AsnGlnCysLeuProAlaLeuGln 228
QY 952 ACCTGCAACAGACAAACCGCACCGCATTTAGTACTAGCCCTCTGCGCCGCAAGCCAC 1011
Db 229 GlnCysSerGlnSerGlySerThr-----SerAspCysThrAsnAlaAsp 243
QY 1012 ACATGTGCGAGGACAAATTTGAGGGCGCATACTAGCCCTTTGCTGGTGGTGTGAT 1071
Db 244 SerValCysTyGlnAsnIleGluGlyProIleSerSerSerGlyAspPheAspValTy 263
QY 1072 GATATTCGCATCATATGATGACCGACTCCGCGCAAGTATTACAAACAAATTTCTGGCA 1131
Db 264 AspIleArgGluProSerAsnAspProTyProProTySerThrTySerThrTyLeuSer 283
QY 1132 AAGGACTCTGTATGACGCTATCGCGCTCAACATCACTACACCGACTCCCAATATGAC 1191
Db 284 AspProThrValVallyAlaIleGlyAlaArgThrAsnTyArgGlnCysProAsnGly 303
QY 1192 GTCTACTACCTTTCAGCAACAGCGGCTTTGTCTGCGCCCACTTCATCGAAGACCTC 1251
Db 304 ProTyAsnLysPheAlaSerThrGlyAspAsnPro---ArgSerPheLeuSerThrLeu 322
QY 1252 GAGGAGATCTTCTCTCCCGTGGGTGCTCTCCCTCATCTATGCGGCGCGGATTCATC 1311
Db 323 SerSerValValGlnSerGlyIleAsnValLeuValTrpAlaGlyAspAlaAspTrpIle 342
QY 1312 TGCAACTGTTCGGCGGTAGCGGCTTTCCTCGTCCGACTACTCCCAAGCGCCAG 1371
Db 343 CysAsnTrpLeuGlyAsnTyGluValAlaAsnAlaValAspPheProGlyAsnAlaGln 362
QY 1372 TTCGGAAGCGGATACAGCCCTCAAGTCAACGGCTCGAGTATGGGAAACATCGC 1431
Db 363 PheSerAlaLeuAspLeuAlaProTyThrValAsnGlyValGluLysGlyGlnPheLys 382

QY 1432 GAGTATGGTAATTTCTCTTCTACTCGCGTCTATGAGGAGGCCATGAAGTCCCATACTAC 1491
Db 383 ThrValAspAsnPheSerPheLeuLysValTyGlyAlaGlyHisGlnValProTyTyTr 402
QY 1492 CAGCCCATCGCTCCCTGCAATTTGTTAAACCGGACTATCTTCGTTGGGATATCGCAGAG 1551
Db 403 GlnProAspThrAlaLeuGlnAlaPheLysGlnIleIle----- 415
QY 1552 GGCAGAGAAGAAG 1563
Db 416 ---GlnLysLys 418
RESULT 2
S51516
C:Species: Absidia zychnae
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text_change 16-Jun-2000
C:Accession: S51516; S78013; S78014
R:Lee, B.R.; Takeuchi, M.; Kobayashi, Y.
Curr. Genet. 27, 159-165, 1995
A:Title: Molecular cloning and sequence analysis of the scpZ gene encoding the serine
A:Reference number: S51516; MUID:95308538; PMID:7788719
A:Accession: S51516
A:Molecule type: DNA
A:Residues: 1-460 <LEB>
A:Cross-references: EMBL:DJ6519; NID:9556466; PIDN:BA03966.1; PID:9995456
A:Experimental source: strain NRIC 1199
A:Accession: S78013
A:Molecule type: protein
A:Residues: 52-62;90-99;367-381 <LEB>
A:Accession: S78014
A:Molecule type: mRNA
A:Residues: 18-460 <LES>
C:Genetics:
A:Gene: scpZ
A:Introns: 64/3; 95/1; 107/2; 138/3; 234/2; 246/1; 270/1; 343/3; 353/2; 384/3; 437/3
C:Superfamily: serine carboxypeptidase
C:Keywords: Glycoprotein; hydrolase; serine carboxypeptidase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-51/Domain: propeptide #status predicted <PRE>
F:52-460/Product: serine-type carboxypeptidase #status experimental <MAT>
F:40-46-119/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:186/Active site: Ser #status predicted
Alignment Scores:
Pred. No.: 9,55e-42 Length: 460
Score: 749.00 Matches: 179
Percent Similarity: 50.10% Conservative: 72
Best Local Similarity: 35.73% Mismatches: 188
Query Match: 24.74% Indels: 62
DB: 1 Gaps: 15
US-09-712-338-1 (1-1668) x S51516 (1-460)
QY 34 TTGGTTCAGCAGTTGGCCCTTCAGGAAGTACACCGCGCTCCGTCGTAGAGACAG 93
Db 18 LeuValCysAlaProValThrValGlnAlaHisProMetSerHisValLeuArgGln 37
QY 94 CTACCCAAAGAACCCCAACCGGGGTCAAGACTTACACCGCAACAAATGTCACCATCCGG 153
Db 38 ---AspGlyAsnAspThr-----SerSerGlyAsnThrThrGlnLeuLys 51
QY 154 TACAAGAACCCCGGGGCGGCTCGAGACTACCCCGGGTGTCAAATCCTACTCT 213
Db 52 TyTrSerPro-----LysLeuCysAsp-----ProAspValLysGlnTySer 66
QY 214 GGTATGTTCAGACACTCTCCCGAGTCCCATACCTCTTCGTGTTTCGAAGCAGACAT 273
Db 67 GlyTyLeuAspAlaAsnAspGluHisTyPhePheTrpPhePheGluSerLysAsn 86
QY 274 AACCCAGAAACTGCACCTATCATTTGTTGATGTTGCTGCGCTGGAAGCGATCTTTG 333

Db 87 AspProLysAsnAspProLeuThrIleTrpLeuAsnGlyGlyProGlyCysSerSerLeu 106
 QY 334 ATCGTCTCTTCGAAGAGTGGCCCTTGGCATGTCATTCGACTTTTGTGACTACATC 393
 Db 107 IleGlyLeuTrpGluGluLeuGlyProCysGlnGlnAsnGlySer-----Ala 122
 QY 394 AACCTCAGTCGTCGACGAGGTCTCCAAATTACTATTCCTGTCCTCCAGCCATTGGAGTC 453
 Db 123 AsnProHisSerTrpHisHisSerSerAsnMetLeuPhePheAspGlnProAspGlyVal 142
 QY 454 GGCTTTTCATATAGTGATACGCTGTGATGGTCCATTAAACCTGTAACTGGGGTTCGTCGAA 513
 Db 143 GlyPheSerTrp-----146
 QY 514 AATTCGAGCTTTCGAGAGTTCAGGCCCGGTACCCCAACCATTCATGCCACTCTGATCGAT 573
 Db 147 -----GlyLysGln-----ThrValSer 152
 QY 574 ACTACCAATCTTCCCGCAGAGCGCTTGGAGATCTCGAAGATTCTTACTGGAATA 633
 Db 153 ThrThrGluAspAlaGluArgAlaTrpThrPheLeuGlnAlaPheTrpGluThrPhe 172
 QY 634 CTAGTCTGGACTCTAGGTGTCAGTCTAAGGACTTCAGTCTATGGACGAGAGCTATGGA 693
 Db 173 Pro-----GlnTrpSerLysLeuAspValHisTrpPheGlyGluSerTrpGly 188
 QY 694 GGGCACTATGGTCTGCATCTTCAATCATTTTTACGACGAGAGTACGAGAAATTCGCAAC 753
 Db 189 GlyHisTrpIleProGlyPheAlaSerHisValAspMetAsnLysLysValGlnSer 208
 QY 754 GGTAGTGTAAAGTGTTCAGCTTAATTTCACTCTCTGGGAATTTAATTCAGGCAATCATC 813
 Db 209 GlyGluGlyGlyValValProLeuLysSerIleGlyValGlyAsnGlyPheIle 228
 QY 814 GAGGAGGCGATCCAGGCCCTTACTACCTCGATTCGCTGTGAACAATACCTACCGTATC 873
 Db 229 AspAlaValIleGlnTrpLysSerTrpProLysMetThrCysHisSerThrTrpProAla 248
 QY 874 AAGGCTGTCAACGACGCGCTTCAACTACATGAAGTTTGCCAAACCAATGCCAATGT 933
 Db 249 ValLeuSerGluGlu-----GluCysAspLysMetGlnGlnIleTrpGluAsnAsp 265
 QY 934 TGCCAGGATTTGATTTCCACCTGCMAACAGACAAACCGCACCGATTCGCTACGACGCC 993
 Db 266 CysLysProAlaAlaGluGlnCysAlaGluSerAsp-----GluAspSer 280
 QY 994 CTCTCGCGCGAAGCCACCAACATGTGCGAGGCAATGTGTAGGGGCCATCTACACGCTTT 1053
 Db 281 AspCysValAsnAlaAsnGlnCys---GlyGlnIleGluGly---IleTrpAlaGln 298
 QY 1054 CTGGTCTGTGTGTATGATATTCGGCATCCATATGATGACCCGACCTCCGCAAGTTAT 1113
 Db 299 SerGlyTrpSerPheTrpAspIleArgGlnGlnGlyAspAsp---ThrProHisProPhe 317
 QY 1114 TACAACAAATTTCTGGCAAGAGCTCTGTCATGAGCGTATCGCGGTCAACATCACTAC 1173
 Db 318 ValAspGlu---LeuAsnLysAlaSerValIleLysGluValGlyAlaArgGlyHisPhe 336
 QY 1174 ACCAGTCCAAATATAGCTTACTACGCTTCCAGCAACACAGCGCTTTGCTGGGCC 1233
 Db 337 SerMetCysSerAspSerValGlyThrAlaPheAlaGlnThrGlyAspGlyAla---Arg 355
 QY 1234 AACTTCATCGAAGCTCGAGAGATCTTCTCTCCCGCGTGTCTCCCTCATCTAT 1293
 Db 356 SerTrpIleProAlaValGluLysLeuLysGluGlyIleProValLeuIleTrpVal 375
 QY 1294 GGCACGCGGATTCATCTGCACTGGTTCGGCGGTGAGCGGTTTCCCTCGTCGCAAC 1353
 Db 376 GlyAspAlaAspValIleCysAsnTrpTrpGlyAsnLeuAspValAlaAspSerLeuLys 395
 QY 1354 TACTCCCAAGCGCCGCTCCGAAGCGCGAGGTACACGCCCTTGAAGTCAACGCGGCTC 1413
 Db 396 TrpAspGlySerAspAlaPheSerLysThrLysLeuGluAlaTrpLysAlaAspGlyLys 415

QY 1414 GAGTATCGGAAACTCCGAGTATGGTAATTCTCTTCACTCCGCTATGAGCGAGGC 1473
 Db 416 GluValGlyGlnPheArgSerAlaAspLysLeuThrPheValArgValTrpGluAlaGly 435
 QY 1474 CATGAAGTCCCACTACACAGCCCATCGCTCCCTCAATGTTTAACCGGACTATCTTC 1533
 Db 436 HisGluValPrometTrpGlnProGluAlaAlaLeuSerMetPheGlnThrTrpIleSer 455
 QY 1534 GGT 1536
 Db 456 Gly 456
 RESULT 3
 T43236
 carboxypeptidase C (EC 3.4.16.5) precursor [validated] - fission yeast (Schizosaccharomyces pombe)
 N;Alternate names: carboxypeptidase Y
 C;Species: Schizosaccharomyces pombe
 C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 08-Sep-2000
 C;Accession: T43236, T37997
 R;Tabuchi, M.; Iwahara, C.; Ohtani, Y.; Ohuchi, N.; Sakurai, J.; Morita, T.; Iwahara, S.
 J. Bacteriol. 179, 4179-4189, 1997
 A;Title: Vacuolar protein sorting in fission yeast: cloning, biosynthesis, transport, and localization of the vacuolar carboxypeptidase Y
 A;Reference number: Z22357; MUID:97352672; PMID:9209031
 A;Accession: T43236
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1002 <TAB>
 A;Cross-references: EMBL:D86560; NID:G3046860; PIDN:BAA25568.1; PID:G3046861
 R;Oliver, K.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1997
 A;Reference number: Z21760
 A;Accession: T37997
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1002 <OLI>
 A;Cross-references: EMBL:Z97209; PIDN:CAB10121.1; GSPDB:GN00066; SPDB:SPAC19G12.10c
 A;Experimental source: strain 972h(-); cosmid c19G12
 C;Comment: The mature carboxypeptidase exists as a heterodimer held together by a disulfide bond.
 C;Genetics:
 A;Gene: cpy1; SPAC19G12.10c
 A;Map position: 1
 C;Superfamily: serine carboxypeptidase
 C;Keywords: disulfide bond; glycoprotein; heterodimer; hydrolase; serine carboxypeptidase
 F;1-18/Domain: signal sequence #status predicted <SIG>
 F;19-521/Domain: propeptide #status predicted <PRO>
 F;200-332/Region: 13-residue repeats (H-H-[EK]-P-G-E-H-W-P-P-P-[MF])
 F;344-425/Region: 9-residue repeats (E-H-H-[RKQ]-G-G-[KE]-[DE]-K)
 F;522-1002/Product: carboxypeptidase C #status experimental <MAT>
 F;627-880/Disulfide bonds: #status predicted
 F;659/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;715,921,978/Active site: Ser, Asp, His #status predicted
 Alignment Scores:
 Pred. No.: 2,71e-35 Length: 1002
 Score: 652.50 Matches: 165
 Percent Similarity: 47.34% Conservative: 66
 Best Local Similarity: 33.81% Mismatches: 174
 Query Match: 21.56% Indels: 83
 DB: 2 Gaps: 15
 US-09-712-338-1 (1-1668) x T43236 (1-1002)
 QY 145 ACCATCCGTCACAGGAACCCGGGCGAGAGCGGTCTCGAGACTACCCCGGTGTCAA 204
 Db 570 ThrLeuArgValLysAspSerLysProGluSerLeuGlyIleAspThr-----ValLys 587
 QY 205 TCTACTCTGGATATGTCGACACTCTCCGAGTCCCATACCTTCTTCTGTTCTTCGAA 264
 Db 588 GlnTrpThrGlyTrpLeuAspValGluAspArgHisLeuPheTrpPheGlu 607
 QY 265 GCCAGACATAACCCAGAACTGCACCTATCACATTGTGGTTGAATGTGGCCCTGGAAGC 324

608 SerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsnGlyGlyProGlyCys 627
 325 GATCTTTGATCGCTCTCGAGAGTGGCGCTTGCATGTCAT---TCGACTTTT 381
 628 SerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerLeuAsnGlyLeu 647
 382 GATGACTACATCAACCTCACTCGTGAACGAGGCTCTCAATTTACTATTCTGTCCAG 441
 648 LysProGluTyrAsnProHisSerTrpAsnSerAsnAlaSerValIlePheLeuAspGln 667
 442 CCATTGGAGTGGCTTTTCATATGATGATACAGGTTCATGGTCCATTACCTGTAAT 501
 668 ProIleAsnThrGlyPheSerAsnGlyAsp-----AspSerValLeuAspThrValThr 685
 502 GGGTCTCGAAATTCAGACTTTTCAGAGGTTTCAGGCGCGGTACCCCAACCATGATGCC 561
 685 ----- 685
 562 ACTCTGATGATACATCAATCTTCCGAGAGCGCTTGGAGATCTCGAAGGATTC 621
 686 -----AlaGlyLysAspValTyrAlaPheLeuAsnLeuPhe 697
 622 CTTAGTGAATCACTAGCTTGGACTCTAGGTGAGCTTAAGGACTTCAGTCTATGAGC 681
 698 PheAlaLysPhePro-----GlnTyrAlaHisLeuAspPheHisIleAlaGly 713
 682 GAGAGCTATGAGGCACTATGCTGCTTCAATCTTTCATGAGCAGAGATGAG 741
 714 GluSerTyrAlaGlyHisTyrIleProGlnPheAlaLysGluIleMetGluHisAsnGln 733
 742 AGA-----ATTGCCAACCGT---AGTGTATGATGTTGTCAGCTTAATTTCAAC 786
 734 GlyAlaAsnPhePheValAlaSerGlyTyrGluMetGluLysGlnTyrIleAsnLeuLys 753
 787 TCTCTGGAAATTAACGATCATCAACGAGGCGATCCAGGCCCTTACTACCTGAA 846
 754 SerValLeuIleGlyAsnGlyLeuThrAspProLeuValGlnTyrTyrPheTyrGlyLys 773
 847 TTCGTGTGAACAATACCTACCGT-----ATCAAG 876
 774 MetAlaCysGluSerProTyrGlyProIleMetSerGlnGluCysAspArgIleThr 793
 877 GCTGTCAACGAGACGCTCTACAACTACATGAAGTTTGCAACCAATGCCAATGGTTGC 936
 794 GlyAlaTyrAspThr-----Cys 799
 937 CAGGATTTGATTTCCACCTGGAACAGACAAACCGACCGCATTAGCTGACTACGCTTC 996
 800 AlaLysLeuIleThrGlyCysTyrGlnThrGlyPheThr-----ProVal 814
 997 TCGCGGAAGCCACCAACATGTGCGAGGCAATGTTGAGGGCCATACCTACGCTTTGCT 1056
 815 CysIleGlyAlaSerLeuTyrCysAsnAsnAlaMetIleGlyPro---PheThrLysThr 833
 1057 GGTCTGTGTGTATGATATCGGCATCCATATGATGACCGGACTCCGCAAGTAT--- 1113
 834 GlyLeuAsnIleTyrAspIleArgGluGluCysArgAspGlnGluHisLeuCysTyrPro 853
 1114 -----TACAACAATTTCCGCAAGGACTCTGTGATGAGCGCTATCGCGCTC 1161
 854 GluThrGlyAlaIleGluSerTyrLeuAsnGlnGluPheValGlnGluAlaLeuGlyVal 873
 1162 AACATCACTACACCCAGTCCAATATAGCTTACTACTAGCTTTCCAGCAAAACAGCGAC 1221
 874 GluTyrAspTyrLysGlyCysAsnThrGluValAsnIleGlyPheLeuPheLysGlyAsp 893
 1222 TTTGTCTGCCCACTTCAAGACCTCGAGAGATCTT-----GCTCTCCCGCTG 1275
 894 TrpMetArgLysThrPheArgAspAspValThrAlaIleLeuGluAlaGlyLeuProVal 913
 1276 CGTGTCTCCTCATCTAT---GGCAGCGCGATTTACATCTGCAACTGGTTCGCGGCTCAG 1332
 914 -----LeuIleTyrAlaGlyAspAlaAspTyrIleCysAsnTyrMetGlyAsnGlu 930

1333 GCGTTTCCCTCGTCCGAACTACTCCCAAGCCCGCCAGTTCGAAAGCGCAGGTACACG 1392
 931 AlaTrpThrAspAlaLeuGluTrpAlaGlyGlnArgGluPheTyrGluAlaGluLeuLys 950
 1393 CCCTGAAAGTCAACCGCGTGCAGTATGCGGAAACTCGCGAGTATGTGTAATTTCTCTTC 1452
 951 ProTrpSerProAsnGlyLysGluAlaGlyArgGlyLysSerPheLysAsnPheGlyTyr 970
 1453 ACTCGGTCTATGAGCGAGCCCATGAAGTCCCATATACCCAGCCCATCGCTCCCTGCAA 1512
 971 LeuArgLeuTyrGluAlaGlyHisMetValProPheAsnGlnProGluAlaSerLeuGlu 990
 1513 TTGTTTAAACCGACTATCTTCGCT 1536
 991 MetLeuAsnSerTrpIleAspGly 998

RESULT 4
 S46008
 Probable carboxypeptidase C (EC 3.4.16.5) YBR139w - yeast (Saccharomyces cerevisiae)
 N;Alternate names: hypothetical protein YBR1015; hypothetical protein YBR139w
 C;Species: Saccharomyces cerevisiae
 C;Date: 26-Aug-1994 #sequence_revision 05-Jan-1996 #text_change 19-Apr-2002
 C;Accession: S46008; S46581
 R;Becam, A.M.; Herbert, C.J.; Nasr, F.; Slonimski, P.P.; Zagulski, M.
 submitted to the Protein Sequence Database, August 1994
 A;Reference number: S45995
 A;Accession: S46008
 A;Molecule type: DNA
 A;Residues: 1-508 <BE>
 A;Cross-references: EMBL:Z36008; NID:G536435; PIDN:CAA85097.1; PID:G536436; GSDB:GN000
 A;Experimental source: strain S288C
 R;Becam, A.M.; Cullin, C.; Grzybowska, E.; Lacroite, F.; Nasr, F.; Ozier-Kalogeropoulos
 yeast 10 (Suppl.A), S1-S11, 1994
 A;Title: The sequence of 29.7kb from the right arm of chromosome II reveals 13 complete
 A;Reference number: S46569; MUID:94378717; PMID:8091856
 A;Accession: S46581
 A;Molecule type: DNA
 A;Residues: 1-508 <BE>
 A;Cross-references: EMBL:X75891; NID:G496856; PIDN:CAA53497.1; PID:G496869
 A;Experimental source: strain S288C
 C;Genetics:
 A;Cross-references: SGD:S0000343
 A;Map position: 2R
 A;Note: MIPS:YBR139w
 C;Superfamily: serine carboxypeptidase
 C;Keywords: hydrolase; serine carboxypeptidase; transmembrane protein
 F.6-22/Domain: transmembrane #status predicted <TMM>
 F.219,415,474/Active site: Ser, Asp, His #status predicted

Alignment Scores:
 Pred. No.: 2,88e-34 Length: 508
 Score: 636.00 Matches: 167
 Percent Similarity: 46.29% Conservative: 70
 Best Local Similarity: 32.62% Mismatches: 197
 Query Match: 21.01% Indels: 78
 DB: 1 Gaps: 14

US-09-712-338-1 (1-1668) x S46008 (1-508)
 QY 55 CTTCCAGGAAGTACACCGCGCTCGGTGAGAGACAGCTA-----CCCAAG 102
 Db 43 LeuProGlnAsnThrGlnGlnThrLeuLysLeuAspArgLeuAsnHisAspProLeu 62
 QY 103 AACCCACCGGGTCAAGACTCTTACACCGCAACAAATGTACCATCCGCTACAGGAA 162
 Db 63 PheThrThrPheIleSerSerValAspThrAspTyrSerLeuArgLeuArgThrValAsp 82
 QY 163 CCGGGGCGAGAGCGGTCTCGGAGACTACCCGGGGTCAAAATCTTCTCTGATATGTC 222
 Db 83 ProSerLysLeuGlyIle-----AspThrValLysGlnTrpSerGlyTyrMet 98
 QY 223 GACACCTCTCCCGAGTCCCATACCTTCTTCTGTTCTTTCGAGCCAGACATACCCAGAA 282


```

QY      1519  ACCCGATATCTTCGGT---TGG 1539
Db      544  AsnArgTrpLeuGlyGlyGluTrp 551
        |||||   :|||
        |||||   :|||

RESULT 6
JCI380
carboxypeptidase C (EC 3.4.16.5) precursor - yeast (Candida albicans)
N;Alternate names: carboxypeptidase Y
C;Species: Candida albicans
C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 24-Nov-1999
C;Accession: JCI380
R;Mukhtar, M.; Logan, D.A.; Kaeufer, N.F.
Gene 121, 173-177, 1992
A;Title: The carboxypeptidase Y-encoding gene from Candida albicans and its
A;Reference number: JCI380; MUID:93051356; PMID:1427093
A;Accession: JCI380
A;Molecule type: DNA
A;Residues: 1-542 <MUK>
A;Cross-references: GB:M95182; MID:g170828; PIDN:AAA34326.1; PID:g170829
A;Note: The authors translated the codon GAT for residue 42 as Asn, AAA for
C;Genetics:
A;Gene: CPY1
C;Superfamily: serine carboxypeptidase
C;Keywords: glycoprotein; hydrolase; serine carboxypeptidase; zymogen
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-127/Domain: propeptide #status predicted <PRO>
F;128-542/Product: carboxypeptidase C #status predicted <VAT>
F;213,291/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;269,461,518/Active site: Ser, Asp, His #status predicted
F;316-330,347-356/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.:          1,29e-31          Length:          542
Score:              596.00           Matches:         169
Percent Similarity: 43.47%           Conservative:    64
Best Local Similarity: 31.53%         Mismatches:     193
Query Match:       19.69%            Indels:         110
DB:                  1               Gaps:           20

US-09-712-338-1 (1-1668) x JCI380 (1-542)

QY      55  CTCCAGGAAGTACACGC-----GGTCGGTCGGTAGAACAGACTACCCAAGAAGACCCC
Db      71  LeuAspGlyLeuThrProGluIleLysAnlleTrpLeuGluMetLeuMetLysPhePro
        |||||   :|||
        |||||   :|||

QY     109  ACCGGGTCAAGACTTTACAACGCCAAACAATGTCAAC---ATCCGGTCAAGAAGAACCC
Db      91  -----AsnSerlleThrGluLeuAsnPheLysAlaPro
        |||||   :|||
        |||||   :|||

QY     166  GGGGCAGAGGGCTTCGCAGACT-----
Db     102  ProLysIleGlyLyslleThrThrGlnGlnPheAspPheHisValThrAspAlaGlnVal
        :|||
        :|||

QY     190  -----ACCCCG-----GGTGTCAAA
Db     122  ProAsnHisiLysLeuargIlleLysSerThrProLysAspLeuGlylleAspThrVallys
        |||||   :|||
        |||||   :|||

QY     205  TCCPACTCTGGATATGFCGACACTCTCCCGAG---TCCCATACCTTCTCTCGGTCTCTTC
Db     142  GlntyrSerGlytyrLeuaspValValaspGluAspLysHisPhePheTyrtyrPhePhe
        |||||   :|||
        |||||   :|||

QY     262  GAAGCCAGACNATAACCCAGAAACTGCACCTATCACATTGTTGTTCAATGGTGGCCCTCGA
Db     162  GluSerArgAsnAspProLysAsnAspProvalIleuTrpPheAsnGlyGlyProgly
        |||||   :|||
        |||||   :|||

QY     322  AGCGATTCTTTGATCGGTCTCTTCGAAAGAGTTGGGCCCTTGCCATGTCAAATPCGACTTTT
Db     192  CysSerSerLeuThrGlyLeuPhePheGluLeuGlyProSerSerIleAspLysAsnLeu
        |||||   :|||
        |||||   :|||

QY     382  GATGACTACATCAACCTCTACTCGTGGAAAGAGGCTTCCAANTTACTATTCTCTGTCGCCAG
Db     202  LysProValtyrAsnProHisSerTrpAsnAlaAsnAlaSerValIlePheLeuAspGln
        |||||   :|||
        |||||   :|||

```



```
QY 442 CCATGGAGTCGGCTTTTCATATAGTATACGGTGTGATGGTCCATTAACCTGTAACT 501
Db 222 ProileAsnValGlyTyrSerTyrSer----- 230
QY 502 GGGTCTCTCGAANAATTCGAGCTTTCAGGAGTTTCAGGGCCGGTACCCAAACCATTTGATGCC 561
Db 231 -----Ser 231
QY 562 ACTCTGATCATACTACCAATCTTTCGCGAGAGCCGCTTGGGAGATCCTGCACAGGATTC 621
Db 232 GlnSerValSerAsnThrIleAlaAlaGlyLysAspValTyrAlaPheLeuGlnLeuPhe 251
QY 622 CTTAGTGGACTACTAGCTTGGACTCTPAGGGTGCAGTCAAGGACTTCAGTCTATGACG 681
Db 252 PheLysAsnPro-----GluTyrAlaAsnLeuAspPheHisIleAlaGly 267
QY 682 GAGAGTATGAGGGCACTATGCTCTGCAATCTTCAATCATTTTACGAGCAGATGAG 741
Db 268 GluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerGluIleLeuThrHisProGlu 287
QY 742 AGAATGCCAACGGTAGTGTAAATGGTTCAGCTTAATTTCACTCTCTGCGAATTAAT 801
Db 288 Arg-----AsnPheAsnLeuThrSerValLeu 296
QY 802 -----AACGGATATCGACGAGCGCATCCAGGCCCTTACTACCTCGAATTCGCTGTG 855
Db 297 IleGlyAsnGlyLeuThrAspProLeuValGlnTyrGluTyrTyrGluProMetAlaCys 316
QY 856 AACAAATACCTACGGTATCAAGGCTGTCACGAGACCGTCTACAACTACATGATGTTGCC 915
Db 317 GlyGluGlyGlyGluProSerValLeuGluProGluGluCysAspGlyMet-----Leu 334
QY 916 AACCAAAATGCCAAATGGTTCGCCAGGATTTGATTTCCACTGCGAAACAGACAAACCCACC 975
Db 335 AsnLeuLeuProArg---CysLeuSerLeuIleGluSerCysTyrGluSer----- 350
QY 976 GCATTAGTCACTACGCCCTCTGCGCCGAGGCAACCAACATGTCGAGGAGCAATGTTGAG 1035
Db 351 -----GlySerValTyrSerCysValProAlaThrIleTyrCysAsnAsnGlyGlnMet 368
QY 1036 GGGCATACTACGCCCTTCTGCTGGTGTGTGTATGATATTCGGCATCCATGATGATAC 1095
Db 369 GlyProTyrGlnLys---ThrGlyArgAsnValTyrAspIleArgThrMetCysGluGly 387
QY 1096 CCG-----ACTCGGCCAAGTATTACAAACAATTTCTGCAAGGACTCTGTC 1143
Db 388 SerSerLeuCysTyrSerGlnLeuGluTyrIleAspGlnTyrLeuAsnLeuProGluVal 407
QY 1144 ATGAGCGCTATCGGCTCAACATCAAC---TACACCCAGTCCCAATAATGACGCTCTACTAC 1200
Db 408 LysLysAlaLeuGlyAlaGluValAspGluTyrGlnSerCysAsnPheAspIleAsnArg 427
QY 1201 GCTTTCAGCAACAGCGACTTGTCTGGCC-----AATTCATGCAAGACCTC 1251
Db 428 AsnPheMetPheAlaGlyAspTrpMetLysProTyrGlnLysAsnValIleAspLeuLeu 447
QY 1252 GAGGAGATCTTGTCTCCCGCTGGTCTCTCCCTCATCTAT---GGGACGCCGATTAAC 1308
Db 448 GluLys-----GluLeuProVal-----LeuIleTyrAlaGlyAspLysAspPhe 462
QY 1309 ATCTGCAACTGGTTCGGCGGTGAGCGCTTTCCTCGCTGGGAATCTCCCAAGCCGCC 1368
Db 463 IleCysAsnTrpLeuGlyAsnGlnAlaTrpThrAsnArgLeuGluTyrSerGlySerLys 482
QY 1369 CAGTTCGAGCGGAGGTACACGCCCTCGAAGTCAACGGCGTGCAGTATGGGAAACT 1428
Db 483 GlyPheThrLysAlaProValLysSerTrpLysValGlyLysAsnAlaAlaGlyGluVal 502
QY 1429 CGCGAGTATGTAATTTCTCTCTCACTCGCTCTATGAGGAGCCCATGAAGTCCCATAC 1488
Db 503 LysAsnTyrLysHisPheThrPheLeuArgValPheGlyGlyGlyHisMetValProTyr 522
```

```
QY 1489 TACCAGCCCATCGCTCCCTGCAATTTGTTTAACCGGACTATCTTCGGT 1536
Db 523 AspGlnProGluAsnAlaLeuAspMetValAsnArgTrpIleSerGly 538
RESULT 7
S61713
carboxypeptidase C (EC 3.4.16.5) precursor - yeast (Pichia pastoris)
N/Alternate names: carboxypeptidase Y
C/Species: Pichia pastoris
C/Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 16-Jun-2000
C/Accession: S61713
R/Ohi, H.; Ohtani, W.; Okazaki, N.; Furuhata, N.; Ohmura, T.
Yeast 12, 31-40, 1996
A/Title: Cloning and characterization of the Pichia pastoris PRC1 gene encoding carboxy-
A/Reference number: S61713; PMID:96381245; PMID:8789258
A/Accession: S61713
A/Molecule type: DNA
A/Residues: 1-523 <OH>
A/Cross-references: EMBL:X87987; NID:g1171615; PIDN:CAA61240.1; PID:g1171616
C/Superfamily: serine carboxypeptidase
C/Keywords: glycoprotein; hydrolase; serine carboxypeptidase; zymogen
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-106/Domain: propeptide #status predicted <PRO>
F;107-522/Product: carboxypeptidase C #status predicted <P>
F;193,271,484,487/Binding site: carboxydrate (Asn) (covalent) #status predicted
F;249,441,498/Active site: Ser, Asp, His #status predicted
F;296-310,327-336/Disulfide bonds: #status predicted
Alignment Scores:
Pred. No.: 1-6e-31 Length: 523
Score: 594.50 Matches: 153
Percent Similarity: 45.89% Conservative: 59
Best Local Similarity: 33.12% Mismatches: 169
Query Match: 19.64% Indels: 81
DB: 1 Gaps: 16
US-09-712-338-1 (1-1668) x S61713 (1-523)
QY 202 AAATCCTACTCTGATATGTCGACACC---TCTCCGAGTCCCATACCTTCTTCTGGTTC 258
Db 121 LysGlnTyrSerGlyTyrLeuAspValGluAlaAspAspLysHisPheTyrTrpPhe 140
QY 259 TTCGAGCCAGACATACCCGAAACCTGACCTATCATCTGTTGGTGAATGGTGGCCCT 318
Db 141 PheGluSerArgAsnAspProGlnAsnAspProIleLeuLeuTyrLeuAsnGlyGlyPro 160
QY 319 GGAACGGATCTCTTCATCGTCTCTTCGAGAGTGGGCCCTTGCATGTCATTCGACT 378
Db 161 GlyCysSerSerLeuThrGlyLeuPheGluLeuGlySerSerArgIleAsnGluAsn 180
QY 379 TTTGATGACTACATCAACCCCTCACTCTGGAACGAGGTCTCCAAATTTACTATTCCTGTC 438
Db 181 LeuLysProIlePheAsnProTyrSerTrpAsnGlyAsnAlaSerIleIleTyrLeuAsp 200
QY 439 CAGCCATTGGAGTCGGCTTTTCATATAGTATGATGATCGGTTGATGGTCCATTAAACCTGTA 498
Db 201 GlnProValAsnValGlyPheSerTyrSer----- 210
QY 499 ACTGGGTGCTCGAAAATTTCGAGCTTTTCGAGAGTTTCAGGCGCCGTACCAACCATTTGAT 558
Db 210 ----- 210
QY 559 GCCACTCTGATCGATACCAATCTTCCGCGAGAGCGCTTGGGAGATCCTCAGGA 618
Db 211 SerSerSerValSerAsnThrValValAlaGlyGluAspValTyrAlaPheLeuGlnLeu 230
QY 619 TTCCTTAGTGACTACCTAGCTAGGACTTAGGTCAGGTCAGGACTTCAGTCTATGG 678
Db 231 PhePheGlnHisPhePro-----GluTyrGlnThrAsnAspPheHisIleAla 246
QY 679 ACGGAGAGCTATGGAGGCACTATGTCCTGCTGCTTCAATCATTTTACGAGCAGAA 738
Db 247 GlyGluSerTyrAlaGlyHisTyrIleProValPheAlaAspGluIleLeuSerGlnLys 266
```


Db 139 LysHisPhePheThrPheGluSerArgAsnAspProAlaLysAspProValIle 158
 QY 298 TTGTGGTTGAATGGTGGCCCTGAAGCGATTCTTTGTATGGTCTCTCGAAGAGTTGGGC 357
 Db 159 LeuTriPLeuAsnGlyGlyProGlyCysSerSerLeuThrGlyLeuPhePheGluLeuGly 178
 QY 358 CCTTGGCATTGCAATTCGACTTTTGAGTACTACATCAACCCCTCACTGCTGGAAAGAGTTC 417
 Db 179 ProSerSerIleGlyProAspLeuLysProIleGlyAsnProTyrSerTyrPheAsnSerAsn 198
 QY 418 TCCAAATTTACTATCTCTGCTCCAGCCATCGGAGTCTGGCTTTTCATATAGTATACGGTT 477
 Db 199 AlaThrValIlePheLeuAspGlnProValAsnValGlyPheSerTyrSer 215
 QY 478 GATGGTCCATTAAACCCCTGTAACCTGGGCTGTCGAAATTCGAGCTTTCGAGGAGTTCAG 537
 Db 216 -----GlySerSer 218
 QY 538 GCGCGTACCCCAACCAATTGATGCCACTGTGATGATPACTACCAATCTTCCCGCAGAGGCC 597
 Db 219 Gly-----ValSerAsnThrValAlaAlaGlyLysAsp 229
 QY 598 GCTTGGGAGACTCTCAAGGATTCCTTAGTGAGTACTAGTGTGAGTCTAGGAGGAG 657
 Db 230 ValTyrAsnPheLeuGluLeuPhePheAspGlnPheProGluTyrValAsnLysGlyGln 249
 QY 658 TCTAAGGACTTCAGTCTATGGACGGAGAGCTATGAGGACCTATGCTCTGCTGATCTTC 717
 Db 250 -----AspPheHisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAla 267
 QY 718 AATCATTTTACGACGAGATGAGAAATGTCACAGGTAGTGTAAATGGTGTTCAGCTT 777
 Db 268 SerGluIleLeuSerHisLysAspArg----- 276
 QY 778 AATTTCAACTCTCTGGGAATATT-----ACGGCATCATCGACGAGCGCATCGAGCC 831
 Db 277 AsnPheAsnLeuThrSerValIleGlyLeuGlyAsnGlyLeuThrAspProLeuThrGlnTyr 296
 QY 832 CCTTACTACCTCGAATTCGCT-----GTGAACAATACCTAC 867
 Db 297 AsnTyrTyrGluProMetAlaCysGlyGluGlyGlyGluProSerValLeuProSerGlu 316
 QY 868 GGTATCAAGGTGTCAACGAGCGCTCTACACTACATGAAGTTTGCCCAACCAATGCCA 927
 Db 317 GluCysSerAlaMetGluAspSerLeu----- 325
 QY 928 AATGTTTCCAGGATTCATTTCCACTCCCAACAGACAAACCGCAGCTAGCTGAC 987
 Db 326 GluArgCysLeuGlyLeuIleGluSerCysTyrAspSerGln-----Ser 340
 QY 988 TACGCCCTCTCGCGGAGCAGCAACCAATGTGACGGGACAATGTTGAGGGGCCATCTAC 1047
 Db 341 ValTyrSerCysValProAlaThrIleTyrCysAsnAsnAlaGlnLeuAlaProTyrGln 360
 QY 1048 GCCTTGTCTGTGTGTGTATGATATTCGGCATCCATCATATGATGAC----- 1095
 Db 361 Arg---ThrGlyArgAsnValTyrAspIleArgLysAspCysGluGlyAsnLeuCys 379
 QY 1096 ---CCGACTCCGCCAAGTTATTACACAAATTTCTGGCAAGGACTCTGTCTAGCAGCT 1152
 Db 380 TyrProThrLeuGlnAsp---IleAspAspTyrLeuAsnGlnAspTyrValLysGluAla 398
 QY 1153 ATCGCGCTCAACATCACTACACCCAGTCC---AATATGACGCTACTACGCTTTCAG 1209
 Db 399 ValGlyAlaGluValAspHisTyrGluSerCysAsnPheAspIleAsnArgAsnPheLeu 418
 QY 1210 CAACAGCGGACTTGTCTGGCCCAACTTCATCGAAGACCTCGAGGAGATCTCTGCTCTC 1269
 Db 419 PheAlaGlyAspTyrMetLysPro---TyrHisThrAlaValThrAspLeuLeuAsnGln 437
 QY 1270 CCCGTGCTGCTCCCTCATCTATGGCAGCGCCGATACATCTGCAACTGTTCGCGGT 1329
 Db 438 AspLeuProIleLeuValTyrAlaGlyAspLysAspPheIleCysAsnTrpLeuGlyAsn 457

QY 1330 CAGGCGCTTTCCCTCCCTCGAAGTACTCCCAAGCCGCCAGTTCGGAAGC----- 1380
 Db 458 LysAlaTrpThrAspValLeuProTyrLysTyrAspGluGluPheAlaSerGlnLysVal 477
 QY 1381 GCAGGGTACAGCCCTCGAAGTCAACGGGCTGAGTATGGGAAACTCGCGAGTATGGT 1440
 Db 478 ArgAsnTrpIleAlaSerIleThrAspGluValAla---GlyGluValLysSerTyrLys 496
 QY 1441 AATTTCTCTCCCTCACTCGCTCTATGAGGAGGCGCATGAATCCCATCTACACAGCCCATC 1500
 Db 497 HisPheThrTyrLeuArgValPheAsnGlyGlyHisMetValProPheAspValProGlu 516
 QY 1501 GCTCCTCCTCAATTTTAAACCGGACTATCTTCGGT 1536
 Db 517 AsnAlaLeuSerMetValAsnGluTrpIleHisGly 528
 RESULT 9
 T03607
 Probable carboxypeptidase C (EC 3.4.16.5) cbp31 - rice
 N:Alternate names: serine-type carboxypeptidase homolog
 C:Species: Oryza sativa (rice)
 C:Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 21-Jul-2000
 C:Accession: T03607
 R:Washio, K.; Ishikawa, K.
 Plant Physiol. 105, 1275-1280, 1994
 A:Title: Organ-specific and hormone-dependent expression of genes for serine carboxypeptidase
 A:Reference number: Z14975; MUID:95062718; PMID:7972496
 A:Accession: T03607
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-429 <WAS>
 A:Cross-references: EMBL:D17587; NID:g409581; PIDN:BAA04511.1; PID:g409582
 A:Experimental source: cv. Yukihikari
 C:Genetics:
 A:Gene: cbp31
 C:Superfamily: serine carboxypeptidase
 C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase
 F76,414,417/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F148,336,393/Active site: Ser, Asp, His #status predicted

Alignment Scores:
 Pred. No.: 1,658-23 Length: 429
 Score: 473.00 Matches: 130
 Percent Similarity: 45.49% Conservative: 77
 Best Local Similarity: 28.57% Mismatches: 178
 Query Match: 15.63% Indels: 70
 DB: 1 Gaps: 16

US-09-712-338-1 (1-1668) x T03607 (1-429)

QY 208 TACTCTGATAT-----GTGACACACCTCTCCGAGTCCCATACCTCTCTCTGTTCTTC 261
 Db 17 HisAlaGlyTyrTyrArgLeuProAsnThrHisAspAlaArgLeuPheTyrPhePhe 36
 QY 262 GAACCCAGACATAACCCA---GAAACTGCACCTATCATTTGTTGTAATGGTGGCCCT 318
 Db 37 GluSerArgGlySerLysGlyGluAspProValIleTyrLeuThrGlyGlyPro 56
 QY 319 GGAAGCGATTTTGTATCGGTCTCTTCGAGAGTGGGCGCTTCGCAATGCAATCGACT 378
 Db 57 GlyCysSerSerGluLeuAlaLeuPheTyrGluAsnGlyProPheHisIleAlaAspAsn 76
 QY 379 TTTGATGACTACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAATTTACTATCTCTGCC 438
 Db 77 MetSerLeuValTrpAsnAspPheGlyTrpAspGlnGluSerAsnLeuIleTyrValAsp 96
 QY 439 CAGCATTGGAGTGGCTTTTTCATATAGTATGATACGGTTGATGGGTCCATTAACCCGTGA 498
 Db 97 GlnProThrGlyThrGlyPheSerTyrSerSerAsnProArgAspThrArgHisAspGlu 116
 QY 499 ACTGGGTCCTCCAAATTCGACTTTCGAGGAGTTCAGGCGGTTCAGGCCGATTCATCA 558

QY	559	GCAC	TCTG	ATCG	ATACTAC	CAAA	TCTTCC	CGCAG	AGCGCGTGTGG	AGATCCTG	CAAG	611
DB	188	-----	-----	-----	-----	-----	---	---	---	---	---	197
QY	619	TTCCT	TAGT	GGAC	TACCTAG	CTTGG	ACTCTAG	GTGC	AGTCTA	AGCACTT	CAGT	678
DB	198	Ph	Pe	He	Ly	Ly	His	Pro	Gl	uP	he	213
QY	679	ACG	GAG	AGC	TATG	GAG	GCCTAT	GCTCTG	CAATCTT	CAATCAT	TTTAC	738
DB	214	Gly	Gl	uSer	Tyr	Ala	Gly	His	Tyr	ile	Pro	233
QY	739	GAG	AGA	ATTG	CCAC	AGGTAG	TGTTA	TGTTG	TTTCA	GTGTTCA	ATTCA	798
DB	234	lys	-----	Ala	Asn	-----	---	---	---	---	---	248
QY	799	ATT	AAC	GC	CA	TCA	TG	ACG	AGG	CGATCC	AGG	858
DB	249	Gly	Asn	Gly	Leu	Thr	Asp	Pro	Ala	ile	in	268
QY	859	AAT	AC	TAC	GGTAT	CA	AGG	TCTCA	AG	ACG	CGTCT	918
DB	269	-----	-----	Met	Asn	Leu	ile	Ly	ser	Asp	Tyr	281
QY	919	CAA	ATG	CCAA	TGTTG	CC	AGG	ATTG	ATTTCC	ACCTG	CAAC	978
DB	282	Lys	Ph	elle	Pro	Pro	Cys	Gl	uP	he	Ala	299
QY	979	TTA	CTG	ACT	AC	TAC	GCCTCTG	CGCG	GAAG	CAAC	ATGTG	1038
DB	300	-----	-----	Ala	ser	Cys	Met	Ala	Ala	Tyr	Met	315
QY	1039	CCAT	ACT	AC	CGCTT	TGCTG	TGCG	TGGTGTG	TATG	ATATTC	CGCAT	1098
DB	316	Ile	Met	Ly	sLeu	Val	Gly	Thr	Ly	sAsn	Tyr	335
QY	1099	ACT	CGC	CAAG	TATTAC	AA	-----	---	---	---	---	1152
DB	336	Leu	Cys	Tyr	Asp	He	ser	Asn	Leu	Gly	uP	355
QY	1153	ATC	GGC	GC	-----	AAC	ATCA	ACTAC	CCCG	TCCAA	TAA	1209
DB	356	Ile	Gly	Val	Gly	Asp	Leu	Gly	uP	he	Val	375
QY	1210	CRA	AC	AGG	CGACTT	TGCTG	GCC	CACTT	CA	TGCA	-----	1266
DB	376	Thr	Asp	-----	---	Trp	Met	Arg	Asn	Leu	Glu	391
QY	1267	CTC	CCG	CGT	GGTGTCT	CCCTC	ATCTAT	TGG	CGCG	CGCG	ATTAC	1326
DB	392	Asp	Gly	ile	Asn	Val	Leu	ile	Tyr	Ala	Gly	411
QY	1327	GGT	CAG	CGCG	TTCCTG	CGTGG	AACTACT	CTCC	AA	CGCG	CCCG	1386
DB	412	Asn	Ser	Arg	Trp	Val	His	Met	Glu	Tyr	Trp	431
QY	1387	TAC	AG	CC	CCCTG	AAAG	CTCA	ACG	CGCTG	AGTATG	GGG	1446
DB	432	Glu	Ser	Pro	Ph	e	Val	Val	Asp	Gly	Ala	451
QY	1447	TC	CTT	CAC	TGCG	TCTAT	GAG	CGAG	CGCC	ATGA	GTC	1506
DB	452	Ser	Ph	e	Leu	Ly	sVal	His	Asn	Ala	Gly	471
QY	1507	CTG	CA	ATTCTT	TAA	CCGG	1524					
DB	472	Leu	Glu	Met	Leu	Arg	477					

RESULT 11

T48977

carboxypeptidase-like protein F14D17.80 [imported] - *Arabidopsis thaliana*
C;Species: *Arabidopsis thaliana* (mouse-ear cress)

```
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text change 24-Oct-2000
```

C;Accession: T48977
F;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.;
submitted to the Protein Sequence Database, April 2000
A;Reference number: 225008
A;Accession: T48977
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-510 <JOR>
A;Cross-references: EMBL:AL353992; GSPDB:GN00061; ATSP:F14D17.80
A;Experimental source: cultivar Columbia; BAC clone F14D17
C;Genetics:

Alignment Scores:

Pred. No.:	3.03e-21	Length:	51
Score:	439.00	Matches:	13
Percent Similarity:	42.43%	Conservative:	83
Best Local Similarity:	25.90%	Mismatches:	19
Query Match:	14.50%	Indels:	98
DB:	2	Gaps:	16

US-09-712-338-1 (1-1668) x T48977 (1-510)

QY	70	CGCGCGTCGTCGGTGAAGACAGCTACCCAGAAACCCACCGGGGTCAAGACTCTTACA	129
Db	82	ProAlaValAlaPArg	88
QY	130	ACCGCAACAATGTCCACATCGGTACAAGNACCGGGGCAGAGGCGCTCTCGAGACT	189
Db	89	-----GlySerGlyGly-----	92
QY	190	ACCGCGGTGTCAAATCCTAC-----TCTGGATAT-----GTGCACACCTCTCCC	234
Db	93	SerProSerValGlnAspPheGlyHisAlaGlyTyrTyrLysLeuProAsnSerLys	112
QY	235	GAGTCCCATACCTTCTTCTGTTCTTCGAGCCAGACATAACCCAGAACTGCACCTATC	294
Db	113	AlaAlaArgMetPheTyrPhePheGluSerArgThrAsn--LysAlaAspProVal	131
QY	295	ACATGTGGTTGAATGGTGGCCCTGGAGCGCATCTTTGATCGGTCCTCTCGAAGAGTTG	354
Db	132	ValIleTrpLeuThrGlyGlyProGlyCysSerSerGluLeuAlaLeuPheTyrGluAsn	151
QY	355	GGCCCTTGGCCATGTCANATGCACTTTTGATGACTACATCAACCCCTCACTCGTGGAAACGAG	414
Db	152	GlyProPheThrValSerAsnAsnSerSerLeuSerTrpAsnGluPheGlyTrpAspLys	171
QY	415	GTCTCCAAATTTACTATTCTCTCCCGCCATGGAGTCGGCTTTTCATATAGTGATACG	474
Db	172	AlaSerAsnLeuIleTyrValAspGlnProValGlyThrGlyPheSerTyrThrSerAsp	191
QY	475	GTTGATGGGTCCATTAAACCTGTAACTGGGGTCGTCGAAAATTCGAGCTTTGCAGGAGTT	534
Db	192	GlnSerAspLeuArgHisAspGlyVal-----	202
QY	535	CAGGGCGGTACCAACCATGTATGCCACTCTGATCGATACTACCAATCTTGCGCCAGAG	594
Db	203	-----SerAsn-----	204
QY	595	GCGCTTTGGAGATCCTGCAAGGATTCTTTAGTGGACTACTAGCTGGACTCTAGGGTG	654
Db	205	AspLeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro-----GlnPhe	220
QY	655	CAGTCTAAGACTTCACTCTATGACGGAGAGCTATGGAGGGCACTATGGTCTCTGCATTC	714
Db	221	ValLysAsnAspPheTyrIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaLeu	240
QY	715	TTCATCATTTTACAGCAGAAATGAGAAATTTGCCAACGGTAGTGTAATGTGTTTCAG	774

241 AlaSerArgValHisArgGlyAsnLys-----AsnLysGluGlyThrHis 255
775 CTTATTTCACTCTCTGGGAATATTAAACGGCATCATCGACGAGGCATCCAGGCCCT 834
256 IleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAsnProGluIleGlnTyrGly 275
835 TACTACCTCGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACGTC 894
276 AlaTyrAlaAspTyrAlaLeuAspMetAsnLeuIleThrGlnSerAspHisAsnLeu 295
895 TACAACCTACATGAAGTTTGCCAACCAAAATGCCAAATGGTTCGACGATTTGATTTCCACC 954
296 AsnArgTyr-----TyrAlaThr 301
955 TCACAACAGACAAACCGCACCGCATTAGCTGACTACGCC-----CTCTGGCCCGAA 1005
302 CysGlnGlnSerIleLysGluCysSerAlaAspGlyGlyGluGlyAspAlaCysAlaSer 321
1006 GCCACCAACATGTGACGAGGACAATGTTTGAGGGGCCATACCTACGCCCTTTGCTGCTGCT 1066
322 SerTyrThrValCys---AsnAsnIlePheGlnLysIleMetAspIleAlaGlyAsnVal 340
1063 GGTGCTGTATGATATTCGGCATCCATATGATGATGACCCGACTCCGCCAAGTATTATTACAAC--- 1119
341 AsnTyrTyrAspValArgLysGlnCysGluGlySerLeuCysTyrAspPheSerAsnMet 360
1120 ---AAATTTCTGCAAGGACTCTGTCATCGGACGCTATCGGGTC---AACTCAACTAC 1173
361 GluAsnPheLeuAsnGlnLysSerValArgLysAlaLeuGlyValGlyAspIleGluPhe 380
1174 ACCAGTCCCAATATGACGCTACTACTACGCTTTCCAGCAACAGCGCACTTGTCTGGCCC 1233
381 ValSerCysSerThrAlaValTyrGluAlaMetGlnMetAsp-----TrpMet 396
1234 AACTTTCATCGAA---GACCTCGAGGAGATCCTTGTCTTCCCTCGGTGTCTCCCTCATC 1290
397 ArgAsnLeuGluValGlyIleProAlaLeuGlnAspGlyIleLysLeuLeuValTyr 416
1291 TATGGGACCGCGATATACATCTGCAACTGTTTGGCGGTCAGGCGCTTCCCTCGCTGCG 1350
417 AlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerLysTyrValHisGluMet 436
1351 AACTTACTCCCAAGCGGCCAGTTCGGAAGCGCGAGGTACACGCGCCCTGAAAGTCAACGGC 1410
437 GluTrpSerGlyGlnLysGluPheValAlaAlaAlaThrValProPheHisValAspAsn 456
1411 GTCGAGTATGGGGAACCTCGCGAGTATGTGAATTTCTCTTCACTCCGCTCATGAGGCA 1470
457 LysGluAlaGlyLeuMetLysAsnTyrGlySerLeuThrPheLeuLysValHisAspAla 476
1471 GGCATGAAGTCCCATACTACCAGCCATCGCTCCCTGCAATGTTTACCGGACTATC 1530
477 GlyHisMetValProMetAspGlnProLysAlaAlaLeuGlnMetLeuGlnAsnTrpMet 496
1531 TTCGGT 1536
497 GlnGly 498
RESULT 12
A35275
carboxypeptidase C (EC 3.4.16.5) - barley
N:Alternate names: carboxypeptidase III
C:Species: Hordeum vulgare (barley)
C:Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 24-Nov-1999
C:Accession: A35275
R:Sorensen, S.B.; Svendsen, I.; Breddam, K.
Carlsberg Res. Commun. 54, 193-202, 1989
A:Title: Primary structure of carboxypeptidase III from malted barley.
A:Reference number: A35275; PMID:90315015; PMID:2639682
A:Accession: A35275
A:Molecule type: protein
A:Residues: 1-411 <SOR>
A:Note: 185-Val was also found

RESULT 12

QY	814	GACGAGCGCATCCAGGCCCTTACTACCTCGAATTGGTGTCGAACAATACCTACCGGTATC	873
		: :	
Dd	181	AspProAlaIle**TyrlysAlatyrThrAspTyrAlaLeuGluMetAsnLeuileGln	200
		: :	
QY	874	AAGCTGTCAACGAGACCCTCTACACTACATGAAGTTTGCACAACCAATGCCAAATGCT	933
		:: :	
Dd	201	LysAlaAspTyrGluArgile-----AsnLysPheIleProPro	213
		: :	
QY	934	TGCAGAGATTGATTTCCACCTGCAACACAGACAAAACGCG- -ACCGCATTTAGCTGACTAC	990
		: :	
Dd	214	CysgluPheAlaallellysleuCysGlyThrAenGlyLysAlaserCysMetalalaIatyr	233
		: :	
QY	991	GCCTCTTCGCGCGAAGCCCAACATGTGCAGGGACAAATGTTGAGGGGGCCNACTACGCC	1050

Db 234 MetValCysAsnThrIlePheAsnSerIleMetLysLeuValGlyThrLysAsnTyr--- 252
 QY 1051 TTGTGTCGTGGTGTCATGATATTCGGCATCCATATGATGACCCGACTCCGCAAGT 1110
 Db 253 -----TyAspValArgLysGluCysGluGlyLysLeuCysTyrAsp 266
 QY 1111 TATTACAAC-----AAATTTCTGCAAAGAGACTCTGTTCATGGACGCTATCGGCGTC--- 1161
 Db 267 PheSerAsnLeuGluLysPhePheGlyAspLysAlaValArgGlnAlaIleGlyValGly 286
 QY 1162 AACATCAACTACACCCAGTCCAAATACACGCTACTACGCTTTCCACAAACACAGCGCAC 1221
 Db 287 AspIleGluPheValSerCysSerThrSerValTyrGlnAlaMetLeuThrAsp----- 304
 QY 1222 TTGTGTCGGCCCACTTCATCGAA---GACCTCGAGGAGATCCTTGTCTCTCCCGTCGGT 1278
 Db 305 -----TrpMetArgAsnLeuGluValGlyIleProAlaLeuLeuGluAspGlyIleAsn 322
 QY 1279 GTCTCCCTCATCTATGCGGACGCGGATTACATTCGCAACTGCTTCGGGGGTACAGCCGTT 1338
 Db 323 ValLeuIleTyrAlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerArgTyr 342
 QY 1339 TCCTCGTCGCGAACTACTCCCAAGCGCCGACGTTCCGAAGCGGAGGTACACGCCCTG 1398
 Db 343 ValHisSerMetGluTrpSerGlyGlnLysAspPheAlaLysThrAlaGluSerSerPhe 362
 QY 1399 AAAGTCAACGGCGTCAGTATGGGAACTCGCGAGTATGTAATTTCTCTTCACTCGC 1458
 Db 363 LeuValAspAlaGlnAlaGlyValLeuLysSerHisGlyAlaLeuSerPheLeuLys 382
 QY 1459 GTCTATGAGCGAGCGCATGAAGTCCCATACTACACGCCCATCGCTCCCTGCAATGTTT 1518
 Db 383 ValHisAsnAlaGlyHisMetValProMetAspGlnProLysAlaAlaLeuGluMetLeu 402
 QY 1519 AACCGG 1524
 Db 403 ArgArg 404
 RESULT 13
 T18968
 Probable serine-type carboxypeptidase (EC 3.4.16.-) Y16B4A.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 31-Mar-2000
 C:Accession: T18968; T23145; T26477
 R:Thomas, K.
 submitted to the EMBL Data Library, March 1996
 A:Reference number: Z19053
 A:Accession: T18968
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2105 <W1>
 A:Cross-references: EMBL:Z70203; PIDN:CAA94110.1; GSPDB:GN000028; CESP:Y16B4A.2
 R:Lloyd, C.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: Z19697
 A:Accession: T23145
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2105 <W1>
 A:Cross-references: EMBL:Z93339; PIDN:CAB07544.1; GSPDB:GN000028; CESP:Y16B4A.2
 R:Wallis, J.
 submitted to the EMBL Data Library, June 1998
 A:Reference number: Z20220
 A:Accession: T26477
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2105 <W1>
 A:Cross-references: EMBL:AL023825; PIDN:CAA19443.1; GSPDB:GN000028; CESP:Y16B4A.2
 A:Experimental source: clone Y16BA4
 C:Comment: This protein has 4 repeats of sequence homologous to serine carboxypeptidase

C:Genetics:
A:Gene: CESP:Y16B4A.2
A:Map position: X
A:Introns: 43/3; 67/2; 179/1; 208/3; 239/2; 281/2; 296/3; 346/2; 444/3; 503/3;
1664/3; 1752/2; 1794/3; 1834/3; 1882/2; 1905/2; 1942/3; 1982/3
C:Keywords: duplication; hydrolase; serine carboxypeptidase

Alignment Scores:
Pred. No.: 1,09e-19 Length: 2105
Score: 417.50 Matches: 156
Percent Similarity: 37.14% Conservative: 91
Best Local Similarity: 23.46% Mismatches: 209
Query Match: 13.79% Indels: 209
DB: 2 Gaps: 28

US-09-712-338-1 (1-1668) x T18968 (1-2105)

```
QY      4  CGTGGCTACGAATTTCTCTCAGTCTACTCCTTGTTCAGCCAGTGGGCC----- 54
Db      1467  GIndGySerGlnTyrValThrThrileAlaGlyTyrAlaLysSerTrpThrGlnAsnLeu 1486
QY      55  -----CTCCAGGAAGTACACCGCGGTCCGTC 81
Db      1487  ValGlnLeuThrValLysGlySerGlyHisPheValProSerAspArgProAlaGln--- 1505
QY      82  GGTAGAAGACAGCTACCCAGAACCCACCGGGGTCAAGACTTTTCAACCGCAACAAT 141
Db      1506  -----AlaLeuGlnMetLeuThrAsnPheLeuSer 1515
QY      142  GTCACCATCCGGTACAGGAACCCGGGACAGGGCGTC----- 180
Db      1516  AsnGlnAlaAsnTyrSerThrProAlaGlyIleAspValThrProGlnProLeuThrAsn 1535
QY      181  -----TGCGAGACT-----ACCCCGGCT 198
Db      1536  ValIleSerGlnThrAsnCysThrThrGlyGlnThrAspArgIleIleAsnLeuProGly 1555
QY      199  GTC-----AAATCCTACTCGNATGTGCACACCTCTCCCGAGTCC 240
Db      1556  LeuProAlaAspMetLeuPheLysGlnTyrSerGlyPheLeuAspGlyLeuSerGlyHis 1575
QY      241  CATACCTCTCTCGTTCCTCGAACCGACACATACCCAGAACTGCACCTATCATGTG 300
Db      1576  LysValHisTyrTrpLeuValGluSerGluAsnAsnProSerThrAspProLeuLeuLeu 1595
QY      301  TGGTTGAATGTGGCCCTGGAAGCATTCTTTGATCGGTCTCTTCGAAGAGTGGGCCCT 360
Db      1596  TrpLeuAsnGlyGlyProGlySerSerLeuMetGlyLeuPheGluGluAsnGlyPro 1615
QY      361  TGCCATGTCAATTCGACTTTTGATGACTACATC---AACCCCTCACTCGTGGAAACAGTCT 417
Db      1616  PheArgValSerLysAspSerGlnThrLeuSerArgAsnProTyrSerTrpAsnLysPhe 1635
QY      418  TCCATTTACTATTCTCTCCAGCCATTGGAGTGGCGCTTTTCATATAGTATACGGTT 477
Db      1636  AlaAsnValLeuTyrLeuGluSerProIleGlyValGlyTyrSerTyrAla----- 1652
QY      478  GATGGTCCATTAAACCCCTGTAACTGGGTCGTCGAAAAATTCGAGCTTTGCAGAGTTCAG 537
Db      1652  ----- 1652
QY      538  GGCCGGTACCAACCATGTGATGCCACTCTGATCGATACTACCAATCTTGGCGCAGAGGCC 597
Db      1653  -----TyrAsnAsnThrAsnIleGlnTyrAspAspValThr-----ThrAlaGlnGlu 1668
QY      598  GCTTGGGAGATCTCGCAAGATTCCTTAGTGGACTCCTAGCTTGGACTCTAGGCTGCAG 657
Db      1669  AsnTyrAlaAlaLeuLysSerPhePheAlaGlnTyrPro-----GlnTyrThr 1684
QY      658  TCTAAGGACTTCAGTCTATCGACGGAGAGCTATGGAGGGCAGCTATGCTCTGCAATCTTC 717
Db      1685  ThrSerAspPheTyrThrThrGlyGluSerTyrAlaGlyValTyrLeuProGlyLeu--- 1703
```

Db 2016 AsnSerAsnPheThrThrThrThrThrSerThrThrThrThrProGlyThrGly 203

Qy 1630 CCTACGGCTACCAGC 1644
||||| ||||| : : : : :
Db 2036 ProThrValThrAla 2040

RESULT 14

A29412
carboxypeptidase C (BC 3.4.16.5) precursor - wheat
N:Alternate names: carboxypeptidase Y homolog
C:Species: Triticum aestivum (common wheat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Nov-1999
J:Accession: A29412
R:Baulcombe, D.C.; Barker, R.F.; Jarvis, M.G.
J. Biol. Chem. 262, 13726-13735, 1987
A:Title: A gibberellin responsive wheat gene has homology to yeast carboxypeptidase C
A:Reference number: A29412; PMID:88007602; PMID:2820978
A:Accession: A29412
A:Molecule type: mRNA
A:Residues: 1-499 <BAU>

C:Genetics:
A:Introns: 100/2; 156/3; 196/3; 244/3; 295/1; 324/3; 411/1; 455/3
C:Superfamily: serine carboxypeptidase
C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase; zymogen
F1-22/Domain: signal sequence #status predicted <SIG>
F23-73/Domain: propeptide #status predicted <PRO>
F74-484/Product: carboxypeptidase C #status predicted <MAT>
F485-499/Domain: carboxyl-terminal propeptide #status predicted <CPRO>
F144/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:216,404,461/Active site: Ser, Asp, His #status predicted

Alignment Scores:

Pred. No.:	9,27e-20	Length:	499
Score:	416.50	Matches:	128
Percent Similarity:	41.39%	Conservative:	81
Best Local Similarity:	25.35%	Mismatches:	221
Query Match:	13.76%	Indels:	75
DB:	1	Gaps:	16

US-09-712-338-1 (1-1668) x A29412 (1-499)

Qy	34 TTGGTTCGACAGTAGTTGGGCCCTTCOCAGGAAGTACCGCGTGCCGTGGTAGAACACAG 93
Db <td>40 LeulleArgAlaLeuAsnLeuLeuProGlyArgProArgArgGlyLeuGly----- 56</td>	40 LeulleArgAlaLeuAsnLeuLeuProGlyArgProArgArgGlyLeuGly----- 56
Qy <td>94 CTACCCCAAGAACCCACCGGGGTCAAGACTCTTACAACGCCAAACAATGFCACCATCCGG 153</td>	94 CTACCCCAAGAACCCACCGGGGTCAAGACTCTTACAACGCCAAACAATGFCACCATCCGG 153
Db <td>57 -----AlaGlyAlaGluAspValalaProGlyGlnLeuLeuGluAaGarg 71</td>	57 -----AlaGlyAlaGluAspValalaProGlyGlnLeuLeuGluAaGarg 71
Qy <td>154 TACAAGGAACCCGGG---GCAGAGGGCGTCTGGGAGACTACCCCGGGTGTCAAATCTCAC 210</td>	154 TACAAGGAACCCGGG---GCAGAGGGCGTCTGGGAGACTACCCCGGGTGTCAAATCTCAC 210
Db <td>72 ValThrLeuProGlyLeuProGlyValGlyAsp-----LeuGlyHisHis 87</td>	72 ValThrLeuProGlyLeuProGlyValGlyAsp-----LeuGlyHisHis 87
Qy <td>211 TCTGGATAT-----GTCGACACCTCTCCGAGTCCCATACCTTCTTGGTCTTCGAA 268</td>	211 TCTGGATAT-----GTCGACACCTCTCCGAGTCCCATACCTTCTTGGTCTTCGAA 268
Db <td>88 AlaGlyTyrTyrArgLeuProAsnThrHisAspAlaArgMetPheTyrPhePheGlu 107</td>	88 AlaGlyTyrTyrArgLeuProAsnThrHisAspAlaArgMetPheTyrPhePheGlu 107
Qy <td>265 GCCAGACATAACCCAGAAAATGCACCTATCACATTGTGTTGAATGTGGCCCTGGAGGC 322</td>	265 GCCAGACATAACCCAGAAAATGCACCTATCACATTGTGTTGAATGTGGCCCTGGAGGC 322
Db <td>108 SerArgGlyLysLysGluAsp---ProValValleItrPheThrGlyGlyProGlyCys 126</td>	108 SerArgGlyLysLysGluAsp---ProValValleItrPheThrGlyGlyProGlyCys 126
Qy <td>325 GATCTTTGTATCGGTCTCTCGAAGAGTGGGCCCTTGCCCATGTCAATTCGATTTTGAT 389</td>	325 GATCTTTGTATCGGTCTCTCGAAGAGTGGGCCCTTGCCCATGTCAATTCGATTTTGAT 389
Db <td>127 SerSerGluLeuAlaValPheTyrGluAsnGlyProPheThrIleAlaAsnMetSer 146</td>	127 SerSerGluLeuAlaValPheTyrGluAsnGlyProPheThrIleAlaAsnMetSer 146
Qy <td>385 GACTACATCAACCCCTCACTCGCTGGAAACAGGTCTCCAATTTACTATTCTCTCCAGCCA 444</td>	385 GACTACATCAACCCCTCACTCGCTGGAAACAGGTCTCCAATTTACTATTCTCTCCAGCCA 444
Db <td>147 LeuValTrpAsnLysPheGlyTrpAspLysIleSerAsnIleIlePheValAspProAla 166</td>	147 LeuValTrpAsnLysPheGlyTrpAspLysIleSerAsnIleIlePheValAspProAla 166
Qy <td>445 TTGGAGATCGGTTTTTCATATATAGTATGATACCGGTGTGATGGTGCATTAAACCTGTACTGGG 501</td>	445 TTGGAGATCGGTTTTTCATATATAGTATGATACCGGTGTGATGGTGCATTAAACCTGTACTGGG 501
Db <td>167 ThrGlyThrGlyPheSerTyrSerSerAspAspArgAspThrArgHisaspGluAlaGly 168</td>	167 ThrGlyThrGlyPheSerTyrSerSerAspAspArgAspThrArgHisaspGluAlaGly 168

QY 505 GTCTGCGAAATTCGAGCTTTGCGAGGATTGAGGCGCGGTACCCAAACCATGTGTCGCACT 564
 Db |||||
 187 Val ----- 187
 QY 565 CTGATCGATGACTACCAATCTTGCGCGCAGAGCCGCTTGGGAGATCTTGCAGAGATTCCTT 624
 Db : : : : : |||||
 188 -----SerAsnAspLeuTyrAspPheLeuGlnValPhePhe 199
 QY 625 AGTGGACTACTAGCTTGGACTCTAGGGTGCAGTCTAAGGACTTCAGTCTATGACGAGGAG 684
 Db |||||
 200 LysLysHisProGluPhe-----ValLysAsnAspPhePheIleThrGlyGlu 215
 QY 685 AGCTATGAGGAGGACTATGCTCTGCAATCTTCAATCATTTTACGAGCAGATGAGAGA 744
 Db |||||
 216 SerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValHisGlnGlyAsnLysLys 235
 QY 745 ATTGCCAACCGTAGTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 804
 Db : : : : : |||||
 236 -----AsnGluGlyThrHisIleAsnLysGlyPheAlaIleGlyAsn 250
 QY 805 GGCATCATGACGAGGCGATGCCAGGCCCTTACTACCTCAATTCGCTGTGAACAATACC 864
 Db |||||
 251 GlyLeuThrAspProAlaIleGlnTyrLysAlaTyrThrAspTyrAlaLeuAsp----- 268
 QY 865 TAGCGTATCAAGCTGTCAACGAGCGCTTCAACTACATGAAGTTTGCACCAACCAATG 924
 Db : : : : : |||||
 269 -----MetAsnLeuIleGlnLysAlaAspTyrAspArgIle-----AsnLysPhe 283
 QY 925 CCAATATGTTGTCAGGATTGTTTCCACCTGCAACACACAAACCGC---ACCGCATTA 981
 Db |||||
 284 IleProProCysGluPheAlaIleLysLeuCysGlyThrAspGlyLysAlaSerCysMet 303
 QY 982 GCTGACTACGCTCTGCCCGGAGCCACCAACATGTGCGAGGACAATGTTGAGGGGCCA 1041
 Db |||||
 304 AlaAlaTyrMetValCysAsnSerIlePheAsnSerIleMetLysLeuValGlyThrLys 323
 QY 1042 TACTACGCTTTGCTGCTGCTGTATGATATTCGCATCCATATGACCCGACT 1101
 Db |||||
 324 AsnTyr-----TyrAspValArgLysGluCysGluGlyLysLeu 336
 QY 1102 CGCCCAAGTATTACAAC-----AAATTTCTGGCAAGGACTCTGTATGAGCGCTATC 1155
 Db : : : : : |||||
 337 CysTyrAspPheSerAsnLeuGluLysPheGlyAspLysAlaValArgGlnAlaIle 356
 QY 1156 GCGGTC---ACATCAACTACACCAGTCCATAATGACGTCTACTAGCTTTCCAGCAA 1212
 Db |||||
 357 GlyValGlyAspIleGluPheValSerCysSerThrSerValTyrGlnAlaMetLeuThr 376
 QY 1213 ACAGCGACTTGTCTGCGCCCAACTTCATCGAA---GACCTCGAGGAGATCTTCTCTC 1269
 Db |||||
 377 Asp-----TrpMetArgAsnLeuGluValGlyIleProAlaLeuLeuGluAsp 392
 QY 1270 CCCGTGGCTGTCTCCTCATCTATGCGGACCGGATATACATCTGCACTGGTTCGCGGT 1329
 Db : : : : : |||||
 393 GlyIleAsnValLeuIleTyrAlaGlyLeuTyrAspLeuIleCysAsnTrpLeuGlyAsn 412
 QY 1330 CAGCGCGTTTCCCTCGCTGCAACTACTCCCAAGCGCCGAGTTCGGAAGCGAGGTAC 1389
 Db : : : : : |||||
 413 SerArgTrpValHisSerMetGluTrpSerGlyGlnLysAspPheAlaLysThrAlaGlu 432
 QY 1390 ACGCCCTGAAGTCAACGGGCTGAGTATGGGAAACTCGCGAGTATGTAATTTCTCC 1449
 Db : : : : : |||||
 433 SerSerPheLeuValAspAspAlaGlnAlaGlyValLeuLysSerHisGlyAlaLeuSer 452
 QY 1450 TTCCTCGCTGTATGAGCGAGGCGCATCAAGTCTCCATACACGAGCCATCGCTCCCTG 1509
 Db : : : : : |||||
 453 PheLeuLysValHisAsnAlaGlyHisMetValProMetAspGlnProLysAlaAlaLeu 472
 QY 1510 CAATTGTTTAAACCGG 1524
 Db : : : : : |||||
 473 GluMetLeuArgArg 477

RESULT 15
 T49079
 serine-type carboxypeptidase like protein - Arabidopsis thaliana
 N;Alternate names: protein F4F15.110
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
 R;Accession: T49079
 R;Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
 submitted to the Protein Sequence Database, April 2000
 A;Reference number: Z25015
 A;Accession: T49079
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-482 <ALC>
 A;Cross-references: EMBL:AL049711; GSPDB:GN00061; ATSP:F4F15.110
 A;Experimental source: cultivar Columbia; BAC clone F4F15
 C;Genetics:
 A;Gene: ATSP:F4F15.110
 A;Map position: 3
 A;Introns: 116/1; 151/1
 C;Superfamily: serine carboxypeptidase

 Alignment Scores:
 Pred. No.: 1-51e-17 Length: 482
 Score: 383.00 Matches: 135
 Percent Similarity: 39.42% Conservative: 68
 Best Local Similarity: 26.21% Mismatches: 202
 Query Match: 12.65% Indels: 110
 DB: 2 Gaps: 18

 US-09-712-338-1 (1-1668) x T49079 (1-482)
 QY 97 CCCAAGAACCCACCGGGTCAAGACTCTTACACCGCAACAACTGCCATCCGGTAC 156
 Db |||||
 32 ProSerProLysArgGlyValSerSerGlyAspThrSerHisPheAsnValIleGln 51
 QY 157 AAGGAA-----CCGGGCGAGGCGCTCTCGAGACTACCCCGGGT----- 198
 Db : : : : : |||||
 52 ArgGluSerValProSerProLysAspLysLeuIleGlnGlnLeuProGlnPro 71
 QY 199 -----GTCAATCTTCTGATATGTCACACCTCTCCGAGTCCCATACC 246
 Db |||||
 72 SerAspValThrPheLysGlnTyrGlyTyrValAlaValAsnLysProAlaGlyArg 91
 QY 247 TTC-----TTCTGGTCTTCGAAAGCCACACATAACCCAGAACTGCACCTACATG 300
 Db : : : : : |||||
 92 PheLeuTyrTyrPheValGluThrIleLysProGlyAsnThrThrProLeuValIle 111
 QY 301 TGGTGTGATGTGGCCCTGGAGGATCTTTGTATCGGTCTCTTCGAAGAGTTGGCCCT 360
 Db |||||
 112 TrpPheAsnGlyGlyProGlyCysSerSerLeuGlyAlaPheLysGluLeuGlyPro 131
 QY 361 TGCCATGTCAATTCGACTTTTGTATGAC---TACATCAACCTCTACTCTGTGAACGAGTC 417
 Db : : : : : |||||
 132 PheArgValHisSerAspGlyLysThrLeuPheArgAsnProTyrSerTrpAsnAsnGlu 151
 QY 418 TCCAAATTTACTATTCCTGTCCCGCAATTCGAGCTTCGCTTTTCATATAGTACGCTT 477
 Db : : : : : |||||
 152 AlaAsnValLeuPheLeuGluThrProValGlyThrGlyPheSerTyrSer----- 168
 QY 478 GATGGGTCCATTAACTGAGTGGGGTCTGCGAAATTCGAGCTTTCGAGGAGTTTCAG 537
 Db |||||
 169 -----AsnSerProIleAsnGlyLysGln 176
 QY 538 GCGCGGTACCAACCATGTATGCCATCTCATCTACTACCAATCTTCGCGAGAGGCC 597
 Db |||||
 177 GlyAspLysAlaThr-----AlaGluAsp 184
 QY 598 GCTTGGGAGATCTCGAAGGATTCCTTAGTGGACTACTAGCTTGGACTTAGGTCGAG 657
 Db : : : : : |||||
 185 AsnTyrMetPheLeuValAsnTrpLeuGluArgPheProGluTyrLysGlyArg----- 202
 QY 658 TCTAAGGACTTCAGTCTATGAGCGGAGAGCTATGGAGGCGCACTATGCTCTGCTATCTTC 717

Mon Nov 24 13:41:04 2003

us-09-712-338-1.rpr

Db 203 ----AspIleTyrIleAlaGlyGlnSerTyrAlaGlyHisTyrValProGlnLeuAla 220
QY 718 AATCATTTTACGAGCAGATGAGAGAAATGCCACGGTAGTGTATATGGTGTTCAGCTT 777
Db 221 GlnIleIleLeuHisArgAsn-----AsnGlnThrLeuIle 232
QY 778 AATTCAACTCTCTGGGAATTATTAACGGCATCATCGACGAGCGATCCAGGCCCTTAC 837
Db 233 AsnLeuArgGlyIleLeuIleGlyAsnProSerLeuAsnArgGluIleGlnAspPhe 252
QY 838 TACCTCAATTGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACGGCTAC 897
Db 253 GlyTyrLysPheMetPhe-----SerHisGlyLeuIleSer--GlnGlnMetAsp 269
QY 898 AACTACATGAAGTTGGCAACCAATGCCAAATGCTGCCAGATTGATTCCACCTGC 957
Db 270 AsnTyrAsnLysPhe-----CysThrAsp-----SerAspLeu 280
QY 958 AACACAGACAACCGCAGCATTAGCTGACTACGCCCTCTGCGCGGAGCCACCAACATG 1017
Db 281 TyrAspTrpAspLysCysHisLeuAlaSerGlnLysIleGluAlaGlnLysThrHisLeu 300
QY 1018 TGCAGGGACAATGTTAGGGGCCATACCTACGCCCTTGTGCTGTGTGTGTATGATATT 1077
Db 301 -----AspIleTyrAsnIle 305
QY 1078 CGCATCCATAT----- 1089
Db 306 TyrAlaProLeuCysLeuAsnSerThrLeuSerSerGluProLysLysCysThrThrIle 325
QY 1090 -----GATGACCCGACTCCGCCAAGTTATTACAAATTTCTGGCAAGGACTCTCTC 1143
Db 326 MetLysAlaAspProCysSerGlyAsnTyrLeuLysAlaTyrLeuAsnIleLysGluVal 345
QY 1144 ATGACGCTATCGCGGTCAAC-----ATCACTACACCCAGTCCCAATAT 1188
Db 346 GlnGluAlaIleHisAlaAsnThrThrLysIleProTyrGluTyrThrSerCysAsnThr 365
QY 1189 GACGCTACTACGCTTCCAGCAACAGCGACTTTGCTGCGCCCAACTTCATCGAAGAC 1248
Db 366 LysLeuLeuTrpGluTrpAsnGluLysAspArgTyrVal-----SerLeuThrProIle 383
QY 1249 CTCGAGCAGATCCTTCTCTCCCGTGGCTGTCTCCTCATCTATGCGCGCAGCCGATTAC 1308
Db 384 LeuGlnGluLeuMetGlyLysGlyValArgValMetLeuTyrAsnGlyAspValAspLeu 403
QY 1309 ATCTGCAACTGTTCGCGGTCCAGCGCTTCCCTGCTGCGGAACACTCTCCCAAGCCGCC 1368
Db 404 ValIleProPheThrSerThrLeuAlaValValLysThrMetAsnLeuThrValValLys 423
QY 1369 CAGTTCGGAAGCAGGATACAGCCCTGAAAGTCAACGGCGTCCAGTATGGGAAACT 1428
Db 424 GluTrpArg-----ProTrpPheThrGlyGlyHisValGlyGlyPheThr 438
QY 1429 CGCGAGTAT--GGAATTCTCTCTCACTCGCGTCTATCAGGCGCCATGAAGTCCCA 1485
Db 439 GluAspTyrLysGlyAsnLeuThrPheValThrValLysGlyAlaGlyHisSerValPro 458
QY 1486 TACTACCGCCATCGCTCCCTGCAATCTTTAACCGGACTATC 1530
Db 459 ThrAspGlnProIleHisAlaLeuAsnIlePheThrSerPheIle 473

Search completed: November 21, 2003, 17:45:50
Job time : 65.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 21, 2003, 16:30:33 ; Search time 28 Seconds
(without alignments)
5602.896 Million cell updates/sec

Title:
Perfect score: 3027
Sequence: 1 atcgcgggtacgaattct.....ccagtggtggtatggcatag 1668

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool_p/US09712338/runat 17112003 170148 10508/app.query.fasta_1.1863
-DB=SwissProt 41 -QPMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFILE=PCO -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CGN 1.1.25 @runat 17112003 170148 10508 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	29.5	423	1	CP31_PENJA
2	652.5	21.6	1002	1	CBP1_SCHPO
3	636	21.0	508	1	YBY9_YEAST
4	599	19.8	542	1	CBP1_CANAL
5	594.5	19.6	523	1	CBP1_PICPA
6	527.5	17.4	532	1	CBP1_YEAST
7	475	15.7	516	1	CBP1_ARYTH
8	473	15.6	429	1	CBP1_ORYSA
9	455	15.0	500	1	CBP1_ORYSA
10	437	14.4	508	1	CBP3_HORVU
11	416.5	13.8	500	1	CBP3_HUMAN
12	360.5	11.9	476	1	CBP1_YEAST
13	355.5	11.7	470	1	YSS2_CABEL
14	353.5	11.7	436	1	CBP2_HORVU
15	351	11.6	482	1	CBP1_NAEFO
16	349	11.5	474	1	PRTP_MOUSE
17	347.5	11.5	516	1	CP23_HORVU
18	334	11.0	480	1	PRTP_HUMAN

19	332	11.0	469	1	YUW5_CABEL	P52717 caenorhabdi
20	331	10.9	729	1	KEX1_YEAST	P09620 saccharomyc
21	326.5	10.8	454	1	YUW6_CABEL	P52715 caenorhabdi
22	318.5	10.5	476	1	CBP2_HORVU	P08818 hordeum vul
23	316	10.4	574	1	YPP3_CABEL	P52716 caenorhabdi
24	312.5	10.3	471	1	VCP_AEDAE	P42660 aedes aegyp
25	312	10.3	510	1	CBP1_ORYSA	P37890 oryza sativ
26	305.5	10.1	423	1	CBP2_WHEAT	P08819 triticum ae
27	297.5	9.8	505	1	YX22_CABEL	P52714 caenorhabdi
28	295.5	9.8	286	1	CBP1_PEA	Q41005 pisum sativ
29	280	9.3	499	1	CBP1_HORVU	P07519 hordeum vul
30	271	9.0	523	1	PEPS_ASPSA	P52719 aspergillus
31	255	8.4	452	1	RISC_MOUSE	Q92045 mus musculu
32	248	8.2	452	1	RISC_RAT	Q92046 rattus norv
33	237	7.8	452	1	RISC_HUMAN	Q92046 rattus norv
34	225.5	7.4	507	1	SXA2_SCHPO	P32825 schizosacch
35	221.5	7.3	531	1	PEPF_ASPNG	P52718 aspergillus
36	208.5	6.9	5703	1	MUSB_HUMAN	Q92045 mus musculu
37	202.5	6.7	5179	1	MUC2_HUMAN	P47179 saccharomyc
38	191.5	6.3	1161	1	DAN4_YEAST	Q02817 homo sapien
39	190	6.3	1367	1	AMYH_YEAST	P08640 saccharomyc
40	167.5	5.5	324	1	CP21_HORVU	P55747 hordeum vul
41	166.5	5.5	1233	1	MUS5_HUMAN	P98088 homo sapien
42	166	5.5	907	1	VGP3_EBV	P03200 Epstein-Bar
43	161	5.3	610	1	MUC4_HUMAN	Q99102 homo sapien
44	161	5.3	860	1	CH12_COCOPO	P54197 coccidioid
45	160.5	5.3	725	1	AGAI_YEAST	P32323 saccharomyc

ALIGNMENTS

RESULT 1						
CP31_PENJA						
ID	CP31_PENJA	STANDARD:	PRT;	423 AA.		
AC	P34946;					
DT	01-FEB-1994 (rel. 28, Created)					
DT	01-FEB-1994 (rel. 28, Last sequence update)					
DT	28-FEB-2003 (rel. 41, Last annotation update)					
DE	Carboxypeptidase S1 (EC 3.4.16.6).					
OS	Penicillium janthinellum (Penicillium vitale).					
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;					
OC	Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.					
OX	NCBI_TaxID=5079;					
RN	[1]					
RP	SEQUENCE.					
RX	MEDLINE=94039747; PubMed=8224168;					
RA	Svensden I., Hofmann T., Endrizzi J., Remington S.J., Breddam K.;					
RT	"The primary structure of carboxypeptidase S1 from Penicillium					
RT	janthinellum";					
RL	FEBS Lett. 333:39-43(1993).					
CC	-!- CATALYTIC ACTIVITY: Preferential release of a C-terminal arginine					
CC	or lysine residue.					
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.					
DR	PIR; S38953; S38953.					
DR	HSSP; P00729; 1YSC.					
DR	MEROPS; S10.008; --					
DR	InterPro; IPR000379; Ser_estr site.					
DR	InterPro; IPR001563; Serine carboxypept.					
DR	Pfam; PF00450; serine carboxypept; 1.					
DR	PRINTS; PR00724; CRBOXPTASEC.					
DR	ProDom; PD001189; Serine_carboxypept; 1.					
DR	PROSITE; PS00131; CARBOXYPEPT SER SER; 1.					
DR	PROSITE; PS00560; CARBOXYPEPT SER HIS; 1.					
KW	Hydrolase; Carboxypeptidase; Glycoprotein.					
FT	ACT SITE 143 143					
FT	ACT SITE 340 340					
FT	ACT SITE 397 397					
FT	ACT SITE 343 343					
FT	BINDING 398 398					
FT	BINDING 398 398					
FT	DISULFID 8 68					
FT	DISULFID 55 300					
FT	DISULFID 223 246					
FT	DISULFID 230 239					

FT CARBOHYD 200 200 N-LINKED (GLNAC...), 423
SQ SEQUENCE 423 AA; 46496 MW; 55F3333608B12729 CRC64;

Alignment Scores: 7.58e-54 Length: 423
Pred. No.: 894.00 Matches: 186
Score: 54.31% Conservative: 66
Percent Similarity: 40.09% Mismatches: 160
Best Local Similarity: 29.53% Indels: 52
Query Match: 29.53% Gaps: 8
DB: 1

US-09-712-338-1 (1-1668) x CPS1_PENJA (1-423)

QY 175 GGCTCTCGGAGATACCCCGGGTGTCAATCTTACTGTGATATGTCACACCTCTCC 234
DB 6 GlyIleCysGluThrThrProGlyValAsnGlnTyrSerGlyTyrLeuSerValGlySer 25
QY 235 GAGTCCCATACCTTCTTCTGTTCTTTCGAAGCCAGACATAACCCAGAACTGCCTATC 294
DB 26 AsnMetAsnMetTyrPheTyrPheGluAlaArgAsnProGlnGlnAlaProLeu 45
QY 295 ACATGTGGTGAATGTGGCCCTGAAGCGATCTTTGATCGTCTCTTTCGAAGAGTTG 354
DB 46 AlaAlaTyrPheAsnGlyGlyProGlyCysSerMetIleGlyLeuPheGlnGluAsn 65
QY 355 GGCCCTTGCCAT--GTCAATTCGACTTTGATGACTATCATCAACCTCACTCGTGGAA 411
DB 66 GlyProCysHisPheValAsnGlyAspSerThrProSerLeuAsnGluAsnSerTyrPhe 85
QY 412 GAGGTCTCCAAATTTACTATCTCTCCAGCAATGGGAGTCGGTCTTTCATATAGTAT 471
DB 86 AsnTyrAlaAsnMetIleTyrIleAspGlnProIleGlyValGlyPheSerTyrGly-- 104
QY 472 ACGGTGTGATGGTCCATTAACCTGTAACTGGGTGCTGCAAAATCGAGCTTTCAGGA 531
DB 104 ----- 104
QY 532 GTTCAGGGCGGTACCCACCATTTGATGCGACTCTGATCGATACCAATCTTCCGCA 591
DB 105 -----ThrAspValThr-----SerThrValThrAlaAla 115
QY 592 GAGGCGCTGGGAGATCTCGAAGATTCCTTAGTGGACTACTAGTGTGAGCTTAGG 651
DB 116 ProTyrValTyrAsnLeuLeuGlnAlaPheTyrAlaGlnArgProGluTyrGluSerArg 135
QY 652 GTGCACTTAAGACTTCTAGTCTATCGGAGAGAGTATCGGAGGCACTATGCTCTGCA 711
DB 136 -----AspPheAlaIlePheThrGluSerTyrGlyGlyHisTyrGlyProGlu 151
QY 712 TTCTTCAATCATTTTACGAGCAGAATGAGAGAATGCCAATGCTAGTGTATGTTGTT 771
DB 152 PheAlaSerTyrIleGluGlnAsnAlaAlaIleLysAlaGlySerValThrGlyGln 171
QY 772 CAGCTTAATTTCAACTCTCTGGGAATATTAAACGGATCATCGACAGGGGATCCAGGC 831
DB 172 AsnValAsnIleValAlaLeuGlyValAsnAsnGlyTyrIleAspSerThrIleGlnGlu 191
QY 832 CCTTACTACCTGAATTCGTGTGCAAAATACCTAGGTATCAAGGTGTCAACAGAAC 891
DB 192 LysAlaTyrIleAspPheSerTyrAsnAsnSerTyrGlnGlnIleLeuAspSerSerThr 211
QY 892 GTCTACACATACATGAGTTGTCACAAATGCTGTCAGGATTTGATTTCC 951
DB 212 ArgAspSerLeuLeuAspAlaTyrAsn-----AsnGlnCysLeuProAlaLeuGln 228
QY 952 ACCTGCAACAGACAAACCGCAGCATAGTCTAGCTACGCCCTCTCGCGCAGGACACC 1011
DB 229 GlnCysSerGlnSerGlySerThr-----SerAspCysThrAsnAlaAsp 243
QY 1012 AACATGTGCGGACATGTTGAGGGCCATACCTAGCGCTTGTGCTGGTGGTGTGTAT 1071
DB 244 SerValCysTyrGlnAsnIleGluGlyProIleSerSerSerSerGlyAspPheAspValTyr 263

RESULT 2

CBPY SCHPO STANDARD; PRT; 1002 AA.
ID CBPY SCHPO
AC O13849; O14366;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase Y precursor (EC 3.4.16.5) (CPY).
GN CPY1 OR PCY1 OR SPAC19G12.10C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., SUBUNIT, SUBCELLULAR LOCATION, AND MUTAGENESIS OF
CYS-627.
RX MEDLINE=97352672; PubMed=9209031;
RA Tabuchi M., Iwaiharu O., Ohtani Y., Ohuchi N., Sakurai J.-I.,
Morita T., Iwaiharu S., Takegawa K.;
RT "Vacuolar protein sorting in fission yeast: cloning, biosynthesis,
transport, and processing of carboxypeptidase Y from
Schizosaccharomyces pombe.";
RL J. Bacteriol. 179:4179-4189(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
Hollroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
Mooney P., Moule S., Jones M., Leather S., McDonald S., McLean J.,
Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

FT	REPEAT	415	423	2-7 (APPROXIMATE)	
FT	ACT_SITE	715	715	BY SIMILARITY.	
FT	ACT_SITE	921	921	BY SIMILARITY.	
FT	ACT_SITE	978	978	BY SIMILARITY.	
FT	BINDING	924	924	SUBSTRATE (BY SIMILARITY).	
FT	BINDING	979	979	SUBSTRATE (BY SIMILARITY).	
FT	DISULFID	627	680	BY SIMILARITY.	
FT	DISULFID	776	789	BY SIMILARITY.	
FT	DISULFID	799	822	BY SIMILARITY.	
FT	DISULFID	806	815	BY SIMILARITY.	
FT	DISULFID	844	851	BY SIMILARITY.	
FT	CARBOHYD	659	659	N-LINKED (GLCNAC...) (POTENTIAL).	
FT	MUTAGEN	627	627	C->T: 38% OF ORIGINAL ACTIVITY.	
SEQ	SEQUENCE	1002 AA;	114237 MW;	4A8D81CFDAB2D854 CRC64;	

Alignment Scores:	
Pred. No.:	3.77e-37
Score:	652.50
Percent Similarity:	47.34%
Best Local Similarity:	33.81%
Query Match:	21.56%
DB:	1

US-09-712-338-1 (1-1668) x CBPY_SCHPO (1-1002)	
QY	145 ACNATCGGTAAAGAACCGGGGACAGAGGCGTTCGAGACTACCCGGGTGTCAA 204
Db	570 ThrLeuArgValLysAspSerLysProGluSerLeuGlyIleAspThr-----Vallys 587
QY	205 TCTACTCTCGATATGTCGACACTCTCCCGAGTCCCATACTCTTCTGGTCTTCGAA 264
Db	588 GluTyrThrGlyTyrLeuAspValGluAspArgHisLeuPheTrpPheGlu 607
QY	265 GCCAGACATAACCCAGAAACGTCACCTATCATCTGTGGTTGAATGGGCGCTCGAGAC 324
Db	608 SerArgAsnAspProGluAsnAspProValLeuTrpLeuAsnGlyGlyProGlyCys 627
QY	325 GATCTTTGATCGGTCTCTTCGAGAGAGTGGGCGCTTGCATGTCAAT---TCGACTTTT 381
Db	628 SerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIleAsnIleGluThrLeu 647
QY	382 GATGACTACATCAACCTCACTGTGGAACGAGGTCTCCATTTACTATTCTGTGCCAG 441
Db	648 LysProGluTyrAsnProHisSerTrpAsnSerAsnAlaSerValIlePheLeuAspGln 667
QY	442 CCATTGGGAGTCGGCTTTTCATATAGTAGATACCGTTGATGGTCCATTAAACCTGTACT 501
Db	668 ProIleAsnThrGlyPheSerAsnGlyAsp-----AspSerValLeuAspThrValThr 685
QY	502 GGGGTGCGTGAATAATTCGAGCTTTGCAGGAGTTTCAGGGCGGTACCCAAACCATTTGATGCC 561
Db	685 ----- 685
QY	562 ACTCTGATGACTACTACCAATCTTGGCGAGAGCGCGCTTGGGAGATCTCTCAAGGATTC 621
Db	686 -----AlaGlyLysAspValTyrAlaPheLeuAsnLeuPhe 697
QY	622 CTTAGTGGACTACCTAGCTTGGACTCTAGGTCGAGTCTAAGGACACTTCAGTCTATGGACG 681
Db	698 PheAlaLysPhePro-----GlnTyrAlaHisLeuAspPheHisIleAlaGly 713
QY	682 GAGAGCTATGAGGGGCACTATGCTCTGCATCTTCCAATCATTTTACGAGCAGAATGAG 741
Db	714 GluSerTyrAlaGlyHisTyrIleProGlnPheAlaLysGluIleMetGluHisAsnGln 733
QY	742 AGA-----ATTGCCAACCGT---AGTGTTAATGGTCTTCAGCTTAAATTTCAAC 786
Db	734 GlyAlaAsnPhePheValAlaSerGlyTyrGluMetGluLysGlnTyrIleAsnLeuLys 753
QY	787 TCTCTGGGAATTATTAAACGGCATCATCGACGAGCGGATCCAGGCCCTTACTACCTGAA 846
Db	754 SerValLeuIleGlyAsnGlyLeuThrAspProLeuValGlnTyrTyrPheTrpGlyIle 773

```
RT 1144 -----TACACAAATTTCTGCGAAGGACTCTGTCATGCGAGCTATCGGGTTC 1161
DB 854 GluThrGlyAlaIleGluSerTyrLeuAsnGlnGluPheValGlnGluAlaLeuGlyVal 873
QY 1162 AACATCAACTACACCACTCCAATAATAGCTTACTACGCTTTCCAGCAACAGCGCAC 1221
DB 874 GluTyrAspTyrLysGlyCysAsnThrGluValAsnIleGlyPheLeuPheLysGlyAsp 893
QY 1222 TTGTCTCGCCCAACTCATCGAAGACTTCGAGGAGATCCCTT-----GCTCTCCCGT 1275
DB 894 TrpMetArgLysThrPheArgAspValThrAlaIleLeuGluAlaGlyLeuProVal 913
QY 1276 CGTGTCTCCCTCATCTAT-----GGCAGCGCGATTCATCTGCACTGCTCGCGGTGAC 1332
DB 914 -----LeuIleTyrAlaGlyAspAlaSerTyrIleCysAsnTyrMetGlyAsnGlu 930
QY 1333 GCCGTTTCCCTCGCTCGCAACTACTCCCAAGCGCGCTTCCAGGCGCGGTACACG 1392
DB 931 AlaTyrThrAspAlaLeuGluThrAlaGlyGlnArgGluPheTyrGluAlaGluLeuLys 950
QY 1393 CCCCTGAAGTCAACGGGCTCAGTATGGGAACCTCGGAGTATGGTAATTTCTCTTC 1452
DB 951 ProTyrSerProAsnGlyLysGluAlaGlyArgGlyLysSerPheLysAsnGlyTyr 970
QY 1453 ACTCGGCTATGAGCGAGCGCATGAGTCCCATATCTACAGCCCATCGCTCCCTCGCAA 1512
DB 971 LeuArgLeuTyrGluAlaGlyHisMetValProPheAsnGlnProGluAlaSerLeuGlu 990
QY 1513 TTGTTTAAACGGGATCTTCGGT 1536
DB 991 MetLeuAsnSerTrpIleAspGly 998

RESULT 3
YBY9_YEAST
ID YBY9_YEAST STANDARD; PRT; 508 AA.
AC P38109;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative serine carboxypeptidase in ESRL-IRAL intergenic region
DE (EC 3.4.16.-).
GN YB139W OR YB1015.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378717; PubMed=8091856;
RA Becam A.-M., Cullin C., Grzybowska E., Lacroite P., Nasr F.,
RA Ozier-Kalogeropoulos O., Palucha A., Slonimski P.P., Zagulski M.,
RA Herbert C.J.;
```

```
RT 1144 -----TACACAAATTTCTGCGAAGGACTCTGTCATGCGAGCTATCGGGTTC 1161
DB 854 GluThrGlyAlaIleGluSerTyrLeuAsnGlnGluPheValGlnGluAlaLeuGlyVal 873
QY 1162 AACATCAACTACACCACTCCAATAATAGCTTACTACGCTTTCCAGCAACAGCGCAC 1221
DB 874 GluTyrAspTyrLysGlyCysAsnThrGluValAsnIleGlyPheLeuPheLysGlyAsp 893
QY 1222 TTGTCTCGCCCAACTCATCGAAGACTTCGAGGAGATCCCTT-----GCTCTCCCGT 1275
DB 894 TrpMetArgLysThrPheArgAspValThrAlaIleLeuGluAlaGlyLeuProVal 913
QY 1276 CGTGTCTCCCTCATCTAT-----GGCAGCGCGATTCATCTGCACTGCTCGCGGTGAC 1332
DB 914 -----LeuIleTyrAlaGlyAspAlaSerTyrIleCysAsnTyrMetGlyAsnGlu 930
QY 1333 GCCGTTTCCCTCGCTCGCAACTACTCCCAAGCGCGCTTCCAGGCGCGGTACACG 1392
DB 931 AlaTyrThrAspAlaLeuGluThrAlaGlyGlnArgGluPheTyrGluAlaGluLeuLys 950
QY 1393 CCCCTGAAGTCAACGGGCTCAGTATGGGAACCTCGGAGTATGGTAATTTCTCTTC 1452
DB 951 ProTyrSerProAsnGlyLysGluAlaGlyArgGlyLysSerPheLysAsnGlyTyr 970
QY 1453 ACTCGGCTATGAGCGAGCGCATGAGTCCCATATCTACAGCCCATCGCTCCCTCGCAA 1512
DB 971 LeuArgLeuTyrGluAlaGlyHisMetValProPheAsnGlnProGluAlaSerLeuGlu 990
QY 1513 TTGTTTAAACGGGATCTTCGGT 1536
DB 991 MetLeuAsnSerTrpIleAspGly 998

Alignment Scores:
Pred. No.: 4,2e-36 Length: 508
Score: 636.00 Matches: 167
Percent Similarity: 46.23% Conservative: 70
Best Local Similarity: 32.62% Mismatches: 197
Query Match: 21.01% Indels: 78
DB: 1 Gaps: 14

US-09-712-338-1 (1-1668) x YBY9_YEAST (1-508)
QY 55 CTTCCAGGAAGTACACCGCGCTCGGTGCGTAGAAGACAGCTA-----CCCAAG 102
DB 43 LeuProGlnAsnThrGlnGlnThrLeuLysLeuAspArgLeuAsnHisAspProLeu 62
QY 103 AACCCACCGGGTCAAGACTCTTACACCGCAACAATGTCCACCATCGGTACAGGAA 162
DB 63 PheThrThrPheIleSerSerValAspThrAspTyrSerLeuArgLeuArgThrValasp 82
QY 163 CCGGGGCGAGAGGGGTCTCGGAGACTACCCCGGGTGCAAAATCTCTCTGATATGTC 222
DB 83 ProSerLysLeuGlyLeu-----AspThrValLysGlnTrpSerGlyTyrMet 98
QY 223 GACACCTCTCCGAGTCCCATACCTTCTCTGTTCTTTCGAGCCGAGACATACCCAGAA 282
DB 99 AspTyrLysAspSerLysHisPhePheTyrTyrPhePheGluSerArgAsnAspProAla 118
QY 283 ACTGCACCTATCATATTGTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 342
DB 119 AsnAspProIleIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 138
QY 343 TTCGAAGAGTGGGCGCTTGGCCATGTCATTCGACTTTTTCGACTTATCATCAACCCCTCAC 402
```

"The sequence of 29.7 kb from the right arm of chromosome II reveals 13 complete open reading frames, of which ten correspond to new genes.";
RL Yeast 10:S1-S11(1994).
[2]
DISCUSSION OF SEQUENCE.
RX MEDLINE=95042830; PubMed=7954890;
RA Nasr F., Becam A.-M., Grzybowska E., Zagulski M., Slonimski P.P., Herbert C.J.;
RT "An analysis of the sequence of part of the right arm of chromosome II of *S. cerevisiae* reveals new genes encoding an amino-acid permease and a carboxypeptidase";
RT Curr. Genet. 26:1-7(1994).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X75891; CAA53497.1; -;
DB EMBL; Z36008; CAA85097.1; -;
DB PIR; S46008; S46008.
DB HSSP; P00729; 1CYP.
DB MEROPS; S10.UPW; -;
DB SGD; S0000343; YBR139W.
DB InterPro; IPR000379; Ser esters site.
DB InterPro; IPR003563; Serine_carbpept.
DB Pfam; PF00450; serine_carbpept; 1.
DB PRINTS; PR00724; CRBOXYPASEC.
DB PRODOM; PD001189; Serine carbpept; 1.
DB PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
DB PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
KW Hypothetical protein; Hydrolase; Carboxypeptidase.
FT ACT SITE 219 219 BY SIMILARITY.
FT ACT SITE 415 415 BY SIMILARITY.
FT ACT SITE 474 474 BY SIMILARITY.
SQ SEQUENCE 508 AA; 57639 MW; AAB2806C8EE2EDE1 CRC64;

139	LeuPheGluLeuGlyProSerSerIleGlyAlaAspMetLysProIleHisAsnProTyr	158
403	TCGTGGAAACGAGGTCCTCCAAATTTACTATTCTCTGCCAGCATTGGGAGTCGGCTTTTCA	462
159	SerTrpAsnAsnAlaSerMetIlePheLeuGluGlnProLeuGlyValGlyPheSer	178
463	TATAGTGATACGGTGTGATGGGTCCATTAAACCTGTAACTGGGGTCGTCAAAATTCGAGC	522
179	TyrGlyAsp	181
523	TTTGCAGGAGTTACGGGCCGTACCCAACCAATTGATGCCACTCTGATCGCATACTACCAAT	582
182	---	188
583	CTTGCCGCAGAGCGCTTGGGAGATCCTGCAAGGATTCTTGTGGACTACCTAGCTTGTG	642
189	LeuAlaGlyLysAspAlaTyrIlePheLeuGluLeuPheGluAlaPheProHisLeu	208
643	GACTCTAGGTGCGAGCTTAAGGACTTCAGTCTATGGACGGAGACTATGGAGGCACTAT	702
209	-----ArgSerAsnAspPheHisIleAlaGlyGluSerTyrAlaGlyHisTyr	224
703	GGTCCTGCATTCCTCAATCATTTTACGAGCAGAAAT---GAGACAAATGCCAACGGTAGT	759
225	IleProGlnIleAlaHisGluIleValValLysAsnProGluArg	239
760	GTTAATAGTGTTCAGCTTAATTTTCAACTCTCTGGGAATATTAAACGGCATCATCGACGAG	819
240	-----ThrPheAsnLeuThrSerValMetIleGlyAsnGlyIleThrAspPro	255
820	GGCATCCAGCCCTTACTACCTCGAATTCGCTGTG---AACATACTACGGTATCAAG	876
256	LeuIleGlnAlaAspTyrTyrGluProMetAlaCysGlyLysGlyTyrHisProVal	275
877	GCTGTCAACGAGACCGTCAACACTACATGAAGTTTCCCAACCAATGCCAAATCGTTGC	936
276	LeuSerSerGluGluCysGluLysMetSerLysAlaAlaGlyArg	290
937	CAGGATTTGATTTCCACTCAACACAGACAAACCGCACCGCATTTAGCTGACTACGCCCTC	996
291	-----CysArgArgLeuAsnLysLeuCystyrAlaSerLysSer	304
997	-----TGCCCGAAGCCACCAACATGTGAGGACAAATGTGAGGGCCCATACTACGCC	1050
305	LeuProCysIleValAlaThrAlaTyrCysAspSerAlaLeuLeuGluProTyrIleAsn	324
1051	TTTGTGGTGGTGTGATGATATTCGCGCATCCATCATATGATGCCCG	1098
325	-----ThrGlyLeuAsnValTyrAspIleArgGlyProCysGluAspAsnSerThrAspGly	343
1099	-----ACTCCGCCCACTTATTCAACAAATTTCTGCAAAAGGACTCTGTCTGATCGAC	1149
344	MetCystyrThrGlyLeuArgTyrValAspGlnTyrMetAsnPheProGluValGlnGlu	363
1150	GCTATCGGCTCAACATC---AACTACACCCAGTCCCAATTAATACGCTCTACTACGTTTC	1206
364	ThrLeuGlySerAspValHisAsnTyrSerGlyCysAspAsnAspValPheThrGlyPhe	383
1207	CAGCAACACGGCATTGTCTGGCCCACTTCATCGAAGACTCGAGGAGATCTTGGT	1266
384	LeuPheThrGlyAspGlySerLysPro---PheGlnGlnTyrIleAlaGluLeuLeuAsn	402
1267	CTCCCGTCGTCTCCCTCATCTATGGCGAGCGCATTCATCATCTGCACTCGTTCGGC	1326
403	HisAsnIleProValIleuIleTyrAlaGlyAspLysAspTyrIleCysAsnTyrLeuGly	422
1327	GTTCAGCGCGTTTCCCTCGCTGCGAACTACTCCCAAGCCCGCCAGTTCCTCAAGCGCAGG	1386
423	AsnHisAlaTrpSerAsnGluLeuGluTyrIleAsnLysArgArgTyrGlnArgArgMet	442
1387	TACACCCCC---CTGAAGTCAAGCGCGTTCGAGTATCGGGAACTCCGAGTAGTGGT	1440

443 LeuArgProTrpValSerLysGluThrGlyGluLeuGlyGlnValLysAsnTyrGly 462
 1441 AATTTCCTCCACTCGGCTCTATGAGCGAGGCGCATCAAGTCCCATCTACCAGCCCATC 1500
 463 ProPheThrPheLeuArgGlyIleTyrAspAlaGlyHisMetValProTyrAspGlnProGlu 482
 1501 GCTCCCTCGCAATGTTTAAACCGGACTATCTTCGGT 1536
 483 AlaSerLeuGluMetValAsnSerTrpIleSerGly 494
 RESULT 4
 ID CBPY CANAL STANDARD; PRT; 542 AA.
 AC P30574;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YSCY).
 CPY1.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 SE SEQUENCE FROM N.A.
 RX MEDLINE=93051356; PubMed=1427093;
 RA Mukhtar M., Logan D.A., Kauffer N.F.;
 RT "the carboxypeptidase Y-encoding gene from Candida albicans and its
 translation during yeast-to-hyphae conversion.";
 RL Gene 121:173-177(1992).
 CC -1- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES.
 CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 broad specificity.
 CC -1- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
 CC -1- INDUCTION: TRANSIENTLY DOWN-REGULATED DURING THE EARLY EVENTS OF
 YEAST TO HYPAE CONVERSION.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; M95182; AAA34326.2; -;
 DR HSSP; P00729; 1CPY.
 DR MEROPS; S10.001; -;
 DR InterPro; IPR000379; Ser esters site.
 DR InterPro; IPR001563; Serine carboxpept.
 DR Pfam; PF00450; serine carboxpept; 1.
 DR PRINTS; PR00724; CRBOXYPTASEC.
 DR ProDom; PD001189; Serine carboxpept; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 21
 FT PROPEP 22 127 POTENTIAL.
 FT CHAIN 128 542 CARBOXYPEPTIDASE Y.
 FT ACT_SITE 269 269 BY SIMILARITY.
 FT ACT_SITE 461 461 BY SIMILARITY.
 FT ACT_SITE 518 518 BY SIMILARITY.
 FT BINDING 464 464 SUBSTRATE (BY SIMILARITY).
 FT BINDING 519 519 SUBSTRATE (BY SIMILARITY).
 FT DISULFID 182 421 BY SIMILARITY.
 FT DISULFID 316 330 BY SIMILARITY.
 FT DISULFID 340 363 BY SIMILARITY.
 FT DISULFID 347 366 BY SIMILARITY.
 FT DISULFID 385 391 BY SIMILARITY.
 FT CARBOHYD 213 213 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 291 291 7FEA6B9F82F9D44F_C8C64.
 SQ SEQUENCE 542 AA; 61044 MW;

Alignment Scores:
 Pred. No.: 1,496-33 Length: 542
 Score: 599.00 Matches: 169
 Percent Similarity: 43.47% Conservative: 64
 Best Local Similarity: 31.53% Mismatches: 193
 Query Match: 19.79% Indels: 110
 DB: 1 Gaps: 20

US-09-712-338-1 (1-1668) x CBPY_CANAL (1-542)

QY 55 CTTCAGAGAGTACACG-----GCTCGTGGTGTGTATGATTCAGTCCAGAACCC 108
 Db 71 LeuAspGlyLeuThrProGluIleLeuAsnIleThrSerGluMetLeuMetLysPhePro 90
 QY 109 ACCGGGTCAAGACTCTTACACCGCAACAAATGTCACC--ATCCGGTCAAGAACCC 165
 Db 91 -----AsnSerIleThrGluLeuAsnPheLysAlaPro 101
 QY 166 GGGCAGAGGCGTCTCGAGACT----- 189
 Db 102 ProLysLysGlyLysIleThrThrGlnPheAspPheHisValThrAspAlaGlnVal 121
 QY 190 -----ACCCG-----GGTGTCAA 204
 Db 122 ProAsnHisLysLeuArgIleLysSerThrProLysAspLeuGlyIleAspThrValLys 141
 QY 205 TCCTACTCTGATATGTCAGACACTCTCCCGAG--TCCCATACCTCTCTCTGCTTCTC 261
 Db 142 GlnTyrSerGlyTyrLeuAspValValAspGluAspLysHisPhePheTyrTyrPhePhe 161
 QY 262 GAAGCCAGACATACCCAGAAACTGCATATCATGTTGGTGAATGGGCGCTGGA 321
 Db 162 GluSerArgAsnProLysAsnAspProValIleLeuThrLeuAsnGlyGlyProGly 181
 QY 322 AGCATCTTTGATCGTCTCTTCCAGAGTGGGCGCTTCCCATGTCATTCGACTTT 381
 Db 182 CysSerSerLeuThrGlyLeuPhePheGluLeuGlyProSerSerIleAspLysAsnLeu 201
 QY 382 GATGACTACATCAACCCCTCACTCGTGAACGAGTCTCCATTTACTATTCTCTGCCAG 441
 Db 202 LysProValTyrAsnProHisSerTrpAsnAlaSerValIlePheLeuAspGln 221
 QY 442 CCATTGGAGTGGCTTTTCATATAGTGAATACGGTGTATGGTTCATTAACCCGTACT 501
 Db 222 ProIleAsnValGlyTyrSerTyrSer----- 230
 QY 502 GGGGTCTGAAAATTCAGCTTTCCAGAGTTCAGGCGGTACCCCAACCATTCATGCC 561
 Db 231 -----Ser 231
 QY 562 ACTCTGATCGATACATCACTATGTCAGAGCGGCTTGGAGATCCTCGAAGATTC 621
 Db 232 GlnSerValSerAsnThrIleAlaGlyLysAspValTyrAlaPheLeuGlnLeuPhe 251
 QY 622 CTTAGTGACTACCTAGCTCTAGGTCAGTCTAAGGTCAGTCTAAGGCTTCAAGTCTATGGAG 681
 Db 252 PheLysAsnPhePro-----GluTyrAlaAsnLeuAspPheHisIleAlaGly 267
 QY 682 GAGAGCTATGAGGCGCATGATGCTCTCATTTCTCAATCATTTTACGAGCAGAATCAG 741
 Db 268 GluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerGluIleLeuThrHisProGlu 287
 QY 742 AGAATGGCAACGGTAGTGTAAAGGTTGAGTGTAAAGTGTAAATTCATCTCTGGAATATT 801
 Db 288 Arg-----AsnPheAsnLeuThrSerValLeu 296
 QY 802 -----AACGCGATCATCAGAGGCGATCCAGGCGCTTACTACCTCAATTCCTGTG 855
 Db 297 IleGlyAsnGlyLeuThrAspProLeuValGlnTyrGluTyrTyrGluProMetAlaCys 316
 QY 856 AACAAATACCTACGGTATCAGGCTGTCAACGAGCGGTCTACAACATCATCATGAAGTTGCC 915

317 GlyGluGlyGlyGluProSerValLeuGluProGluLeuCysAspGlyMet-----Leu 334
 QY 916 AACCAATGCCAATGGTTGCCAGATTGATTCACCTGCAACAGACAAACCGCAC 975
 Db 335 AsnSerLeuProArg--CysLeuSerLeuIleGluSerCysTyrGluSer----- 350
 QY 976 GCATTAGTACTACGCCCTCTCGCGGAGCAGCACCACCAATGTCGAGGACAATGTTGAG 1035
 Db 351 -----GlySerValTrpSerCysValProAlaThrIleTyrCysAsnGlyGlnMet 369
 QY 1036 GGGCCATACATACGCTTTGCTGCTGTGTGTATGATTCGTCATCGCATCATGATGAC 1095
 Db 369 GlyProTyrGlnLys---ThrGlyArgAsnValTyrAspIleArgThrMetCysGluGly 387
 QY 1096 CCG-----ACTCCGCCCAAGTATTATCAACAAATTTCTGGCAAGGACTCTGTC 1143
 Db 388 SerSerLeuCysTyrSerGlnLeuGluTyrIleAspGlnTyrLeuAsnLeuProGluVal 407
 QY 1144 ATGACGCTATCGCGTCAACATCAAC--TACACCCAGTCCCAATATGAGTCTACTAC 1200
 Db 408 LysLysAlaLeuGlyAlaGluValAspGluTyrGlnSerCysAsnPheAspIleAsnArg 427
 QY 1201 GCTTTCAGCAACAGCGGACTTTGTCTGGCCC-----AACTTCATCGAAGACCTC 1251
 Db 428 AsnPheMetPheAlaGlyAspTrpMetLysProTyrGlnLysAsnValIleAspLeuLeu 447
 QY 1252 GAGAGATCTCTGCTCTCCCGCGTGTCTCTCTCTCTCTAT---GGCAGCGCATATC 1308
 Db 448 GluLys-----GluLeuProVal-----LeuIleTyrAlaGlyAspLysAspPhe 462
 QY 1309 ATCTGCAACTGTTGCGGCGTTCAGCGGTTTCCCTCTCGCACTACTCCCAAGCGGCC 1368
 Db 463 IleCysAsnTrpLeuGlyAsnGlnAlaTrpThrAsnArgLeuGluTrpSerGlySerLys 482
 QY 1369 CAGTTCGAAAGCGGAGGTACACGCCCTCGAAAGTCAACGCGTCGAGTATGGGAAACT 1428
 Db 483 GlyPheThrLysAlaProValLysSerTrpLysValGlyLysAsnAlaAlaGlyGluVal 502
 QY 1429 GCGAGATGATGTAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1488
 Db 503 LysAsnTyrLysHisPheThrPheLeuArgValPheGlyGlyGlyHisMetValProTyr 522
 QY 1489 TACAGCCCATCGCT 1536
 Db 523 AspGlnProGluAsnAlaLeuAspMetValAsnArgTrpIleSerGly 538

RESULT 5
 ID CBPY_PICPA STANDARD; PRT; 523 AA.
 AC PS2710;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YSCY).
 GN PC1.
 OS Pichia pastoris (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4922;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 108-142.
 RC STRAIN=GT5115;
 RX MEDLINE=96381245; PubMed=8789258;
 RA Ohl H., Ohtani W., Okazaki N., Furuhata N., Ohmura T.;
 RT "Cloning and characterization of the Pichia pastoris PRCl gene
 encoding carboxypeptidase Y.";
 RL Yeast 12:31-40(1996).
 CC -- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES.
 CC -- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 CC broad specificity.
 CC -- SUBCELLULAR LOCATION: LYOSOME-LIKE VACUOLES.
 CC -- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.

GOLGI-MODIFIED P2 PRECURSOR (69 kDa). P2 IS SORTED AWAY FROM SECRETORY PROTEINS AT OR BEYOND A LATE GOLGI COMPARTMENT AND IS SUBSEQUENTLY DELIVERED TO THE VACUOLE VIA A PREVACUOLAR ENDOSOME-LIKE COMPLEMENT. UPON ARRIVAL IN THE VACUOLE, THE N-TERMINAL PROSEGMENT OF P2 IS CLEAVED TO YIELD THE ENZYMATICALLY ACTIVE MATURE VACUOLAR FORM OF CPY (61 kDa).

-! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.

-! DATABASE: NAME=worthington enzyme manual; WWW="http://www.worthington-biochem.com/manual/C/COY.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to licenses@isb-sib.ch).

EMBL; M15482; AAA34902.1; --
EMBL; X80836; CAA56806.1; --
PIR; A26597; CPBYV.
PDB; LYSC; 22-JUN-94.
PDB; ICPY; 15-SEP-95.
MEOPS; S10.001; --
SGD; S0004912; PRCI.
GO; GO:0005783; C:endoplasmic reticulum; IDA.
InterPro; IPR000379; Ser esters site.
InterPro; IPR001563; Serine_catpept.
Pfam; PF00450; serine catpept; 1.
PRINTS; PR00724; CRBOXYPTASEC.
ProDom; PD001189; Serine catpept; 1.
PROSITE; PS00131; CARBOXYPEPT SER_SBR; 1.
PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal;
3D-structure.

KW SIGNAL 1 20
FT PROPEP 21 111 POTENTIAL.
FT CHAIN 112 532 CARBOXYPEPTIDASE Y.
FT ACT_SITE 257 257 BY SIMILARITY.
FT ACT_SITE 449 449
FT ACT_SITE 508 508 SUBSTRATE.
FT BINDING 452 452 SUBSTRATE.
FT BINDING 509 509
FT DISULFID 167 409
FT DISULFID 304 318
FT DISULFID 328 351
FT DISULFID 335 344
FT DISULFID 373 379
FT CARBOHYD 124 124
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .)
FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .)
FT CARBOHYD 479 479 N-LINKED (GLCNAC. . .)
FT SITE 24 27 VACUOLAR TARGETING SIGNAL.
FT MUTAGEN 508 508 H->A,R: INACTIVATES ENZYME.
FT CONFLICT 260 261 GH -> HG (IN REF. 3).
FT CONFLICT 389 389 Y -> E (IN REF. 3).
FT CONFLICT 529 529 G -> D (IN REF. 3).
FT STRAND 114 114
FT HELIX 116 118
FT STRAND 129 134
FT TURN 135 138
FT STRAND 139 146
FT TURN 152 154
FT STRAND 157 161
FT TURN 164 166
FT STRAND 167 167
FT TURN 169 169
FT HELIX 170 173
FT TURN 174 177
FT STRAND 180 183
FT TURN 184 186
FT STRAND 187 190
FT TURN 192 193

P00729;
21-JUL-1996 (Rel. 01, Created)
01-NOV-1988 (Rel. 09, Last sequence update)
15-SEP-2003 (Rel. 42, Last annotation update)
Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YSCY).
PCYL OR YMR297W.
Saccharomycetes cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
[1] SEQUENCE FROM N.A. PubMed=3028649;
MEDLINE=87131100; PUBMED=3028649;
Vallis L.A., Hunter C.P., Rothman J.H., Stevens T.H.;
"Protein sorting in yeast: the localization determinant of yeast vacuolar carboxypeptidase Y resides in the propeptide."; Cell 48:887-897(1987).
[2] SEQUENCE FROM N.A.
STRAIN=S288c / AB972;
PubMed=9169872;
Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T., Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S., Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
"The nucleotide sequence of Saccharomyces cerevisiae chromosome XIII."; Nature 387:90-93(1997).
[3] SEQUENCE OF 112-532.
Svendsen I., Martin B.M., Viswanatha T., Johansen J.T.;
"Amino acid sequence of carboxypeptidase Y. II. Peptides from enzymatic cleavages."; Carlsberg Res. Commun. 47:15-27(1982).
[4] REVISIONS, AND ACTIVE SITE SER-257.
Breddam K., Svendsen I.;
"Identification of methionyl and cysteinyl residues in the substrate binding site of carboxypeptidase Y."; Carlsberg Res. Commun. 49:639-645(1984).
[5] ACTIVE SITE HIS-508.
MEDLINE=90315013; PUBMED=2639680;
Bech L.M., Breddam K.;
"Inactivation of carboxypeptidase Y by mutational removal of the putative essential histidyl residue."; Carlsberg Res. Commun. 54:165-171(1989).
[6] MUTAGENESIS.
MEDLINE=94114535; PUBMED=7904479;
Mortensen U.H., Remington S.J., Breddam K.;
"Site-directed mutagenesis on (serine) carboxypeptidase Y. A hydrogen bond network stabilizes the transition state by interaction with the C-terminal carboxylate group of the substrate."; Biochemistry 33:508-517(1994).
[7] X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
MEDLINE=95244421; PUBMED=7727362;
Endrizzi J.A., Breddam K., Remington S.J.;
"2.8-A structure of yeast serine carboxypeptidase."; Biochemistry 33:1106-1120(1994).
-! FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES. DIGESTS PREFERENTIALLY PEPTIDES CONTAINING AN ALIPHATIC OR HYDROPHOBIC RESIDUE IN P1' POSITION, AS WELL AS METHIONINE, LEUCINE OR PHENYLALANINE IN P1 POSITION OF ESTER SUBSTRATE.
-! CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a broad specificity.
-! ENZYME REGULATION: INHIBITED BY ZPCK.
-! SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
-! PTM: ENTERS THE ENDOPLASMIC RETICULUM AS AN INACTIVE ZYMOMEN AND IS MODIFIED BY FOUR N-LINKED CORE OLIGOSACCHARIDES, GIVING RISE TO A PRECURSOR KNOWN AS P1 (67 kDa). AS P1 TRANSITS THROUGH THE

QY 1270 CCCTGCTGCTCCCTCATCTATGCGGACGCGGATTACATCTGCAACTGGTTCGGCGGT 1329
 Db 438 AspleuProilleuValTyAlaGlyAspPheileCysAsnTrpLeuGlyAsn 457
 QY 1330 CAGCCGCTTCCCTCGCTGCGAACTACTCCCAAGCCGCCAGTTCGGAAGC----- 1380
 Db 458 LysAlaTrpThrAspValLeuProTyrLysTyAspGluGluPheAlaSerGlnLysVal 477
 QY 1381 CGAGGTACAGCCCTGAAAGTCAACGGGTCGAGTATGGGAACTCCGAGTATGGT 1440
 Db 478 ArgAsnTrpThrAlaSerIleThrAspGluValAla---GlyGluValLysSerTyLys 496
 QY 1441 AATTCTCTCTCACTCGCTGCTATGAGGAGCGCATGAGTCCCATATACAGCCCATC 1500
 Db 497 HisPheThrTyLeuArgValPheAsnGlyHisMetValProPheAspValProGlu 516
 QY 1501 GCCTCCCTCGCAATGTTTAACCGGAGTATCTTCGGT 1536
 Db 517 AsnAlaLeuSerMetValAsnGluTrpIleHisGly 528

RESULT 7

ID CBPX ARATH STANDARD; PRT; 516 AA.
 AC P32826; Q42107; Q9CAE5;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Serine carboxypeptidase precursor (EC 3.4.16.-).
 GN AT3G10410 OR F13M14.32.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TAXID=3702;
 RN [1]
 RP Bradley D.;
 RL Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RC MEDLINE=21016720; PubMed=11130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,
 RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Delseny M., Bonty M., Grievell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choisine N., Artiguenave F., Robert C., Brottier P.,
 RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Drzonek H., Erle H., Jordan R., Bangert S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
 RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laude M., Berger-Llauro C., Purnelle B., Masuy D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
 RA Manthaupt G., Haase D., Schoof H., Rued S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Creney T.H., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Ideawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana";
 RL Nature 408:820-822 (2000).
 RN [3]

RP SEQUENCE OF 252-372 AND 455-516 FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Seedling;
 RA Hofte H.;
 RL Submitted (SEP-1993) to the EMBL/GenBank/DDBJ databases.
 CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 CC broad specificity
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M81130; AAB04606.1; --
 DR EMBL; AC011560; AAG51389.1; --
 DR EMBL; Z25955; CA81115.1; --
 DR EMBL; Z26528; CA81299.1; --
 DR HSSP; P00729; LYSC.
 DR MEROPS; S10.009; --
 DR InterPro; IPR000379; Ser esters site.
 DR InterPro; IPR001563; Serine carbpept.
 DR Pfam; PF00450; serine_carbpept; 1.
 DR PRINTS; PR00724; CRHOXYPTASEC.
 DR PRODom; PD001189; Serine carbpept; 1.
 DR PROSITE; PS00131; CARBOXYPEPT SER SER; 1.
 DR PROSITE; PS00560; CARBOXYPEPT SER HIS; 1.
 KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; signal.
 FT SIGNAL 1 22
 FT PROPEP 23 82
 FT CHAIN 83 516
 FT ACT_SITE 229 229
 FT ACT_SITE 417 417
 FT ACT_SITE 474 474
 FT BINDING 420 420
 FT CARBOHYD 157 157
 FT DISULFID 139 379
 FT DISULFID 300 322
 FT DISULFID 307 315
 FT CONFLICT 515 516
 FT QM -> ADVITSSPALKHKKQIIKQCVSN (IN REF.
 FT 1).
 SQ SEQUENCE 516 AA; 57301 MW; 177C778DF657A1C1 CRC64;
 Alignment Scores:
 Pred. No.: 4,83e-25 Length: 516
 Score: 475.00 Matches: 150
 Percent Similarity: 43.11% Conservative: 72
 Best Local Similarity: 29.13% Mismatches: 187
 Query Match: 15.63% Indels: 106
 DB: 1 Gaps: 20
 US-09-712-338-1 (1-1668) x CBPX_ARATH (1-516)
 QY 28 CTACCTTCGTTCGACCCAGTTCGGCCCTTCAGGAGTACACCGCGTCGCTCGTAGA 87
 Db 66 LeuProLeuThrAlaAla-----GluGlyProGlyIleValGluA9 79
 QY 88 AGACAGCTACCCCAAGAAC-----CCACCGGGGTCAAGACTCTTACACCGCAACAAT 141
 Db 80 LysPheValPheProAsnIleLeuAlaaspGlyGlyProThrValaspAspLeuGlyHis 99
 QY 142 GTCACATCCCGGTACAGGAACCGGGGAGAGGGCGTCTGCAGACTACCCCGGGTGC 201
 Db 100 HisAlaGlyTyTyTyLysLeuProLysSerArgly----- 111
 QY 202 AATCTCTACTCTGGATATGCGACACTCTCCGAGTCCCATACCTTCTTCTGGTCTTC 261
 Db 112 -----AlaSerMetPheTyPhePhePhe 119
 QY 262 GAAGCCAGACATAACCCAGAACTGCACCTATCATATGTTGGTTGAATGGTGGCCCTGGA 321
 |||:||||| ||| ||| |||:||||| |||:||||| |||:||||| |||:|||||

Db 120 GluSerArg---AsnLysLysAspAlaProValValIleTrpLeuThrGlyGlyProGly 138
 QY 322 AGCAATCTTTGATCGCTCTTCGAGAGAGTGGCCCTTCGCCATGCAATTCGACATTTT 381
 Db 139 CysSerSerGluLeuAlaValPheTyrGluAsnGlyProPheLysIleThrSerAsnMet 158
 QY 382 GATGACTACATCAACCTCTACTCGTGGAAAGAGGTCTCCAAATTTACTATTCTCTGCCAG 441
 Db 159 SerLeuAlaTrpAsnGluTyrGlyTyrAspGlnValSerAsnLeuLeuTyrValAspGln 178
 QY 442 CCATGGGAGTGGCTTTTCATATAGTAGTACGGTTCATGGTCCATTACCCCTGTAAC 501
 Db 179 ProValGlyThrGlyPheSerTyrThrAspLysSerAspIleArgHisAspGluThr 198
 QY 502 GGGGTCTCGAAAATTTCGAGCTTTTCGAGGAGTTTCAGGGCGGTACCCCAACATTGATGCC 561
 Db 199 GlyVal----- 200
 QY 562 ACTCTGATCGATCTACTACCAATCTTGGCGCAGAGCGCTTTGGAGATCCTGCAAGATTC 621
 Db 201 -----SerAsnAspLeuTyrAspPheLeuGlnAlaPhe 211
 QY 622 CTTAGTGGACTACTAGTCTGACTCTAGGTCGACTCTAAGGACTCTAGTCTATGAGC 691
 Db 212 PheAlaGluHisProLysLeu-----AlaLysAsnAspPheTyrIleThrGly 227
 QY 692 GAGAGCTATGAGGGCACTATGCTCTGCAATCTTCAATCAATTTTACGAGCAGATGAG 741
 Db 228 GluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValHisLysGlyAsnLys 247
 QY 742 AGAATGCCAACGGTAGTGTAAATGTTCTAGCTTAATTTCACTCTCTGGAATATT 801
 Db 248 -----AlaAsn-----GluGlyValHisIleAsnLeuLysGlyPheAlaIleGly 262
 QY 802 AACGGCATCATCGAGCGGATCCAGGCCCTTACTACCTGCAATTCCTGTAACAAT 861
 Db 263 AsnGlyLeuThrAspProAlaLeuGlnTyrProAlaTyrProAspTyrAlaLeuGlu--- 281
 QY 862 ACTACGGTATCAAGCTCTCAACGAGACCGCTCTACAATCTACATGAAGTTTGCCACCAA 921
 Db 282 ---MetGlyLeuIleThrGlnLysGlu-----HisAspArgLeuGluLysIle 296
 QY 922 ATGCCAAATGTTGCGAGATTTGATTTCCACTGCAACACAGACAAACCGCACCGCA--- 978
 Db 297 ValProLeu---CysGluLeuSerIleLysLeuCysGlyThrAspGlyThrThrSerCys 315
 QY 979 TTAGCTGACTACGCCCTCTGCGCCGAGCCCAACATGTCGAGGACAATGTTGAGGG 1038
 Db 316 LeuAlaSerTyrLeuValCysAsnSerLeuPheSerGlyValMetSerHisAlaGlyGly 335
 QY 1039 CCACTACTACGCCCTTTGCTGCTGCTGTGTATGATATTGGCATCCA----- 1086
 Db 336 ValAsnTyr-----TyrAspIleArgLysLysCysValGlySer 348
 QY 1087 -----TATGATGACCGACTCCGCCCAAGTATTAACAATAATTTCTGGCAAGACTCT 1140
 Db 349 LeuCysTyrAspPhe-----SerAsnMetGluLysPheLeuAsnLeuGlnSer 364
 QY 1141 GTCATGGAGCTATCGCGGTC---AACATCACTACACCCAGTCCAAATGACGCTTAC 1197
 Db 365 ValArgLysSerLeuGlyValGlyAspIleAspPheValSerCysSerThrSerValTyr 384
 QY 1198 TACGCTTTCCAGCAACAGCGGACTTTGCTGGCCCACTTCATCGAAGACTCGAGGAG 1257
 Db 385 GlnAla-----MetLeuValAspTrpMetArgAsnLeuGlu--- 396
 QY 1258 ATCCTGTCTCTCCCGTGGTGTCTCC-----CTCATCTAT---GGCGAC 1299
 Db 397 ---ValGlyIleProThrLeuLeuGluAspGlyIleSerLeuLeuValTyrAlaGlyGlu 415
 QY 1300 GCCGATTACATCTGCACTGTTGCGCGGTGAGCCGCTTCCTCGCTGCGCACTAC 1359
 Db 416 TyrAspLeuIleCysAsnTrpLeuGlyAsnSerArgTrpValAsnAlaMetGluTrpSer 435

QY 1360 CAAGCCGCCAGTTCGAGAGCGGAGGTACACGCCCTCGAAAGTCAACGGCGTTCGAGTAT 1419
 Db 436 GlyLysThrAsnPheGlyAlaAlaLysGluValProPheIleValAspGlyLysGluAla 455
 QY 1420 GGGGAAACTCGCGAGTATGTAATTTCTCCTTCACTCGCGCTCTATGAGGAGCGCATGAA 1479
 Db 456 GlyLeuLysThrTyrGluGlnLeuSerPheLeuLysValArgAspAlaGlyHisMet 475
 QY 1480 GTCCCATACTACCGACCCCATCGCTCCCTGCAATTTGTTAAACCGG 1524
 Db 476 ValProMetAspGlnProLysAlaAlaLeuLysMetLeuLysArg 490

RESULT 8

CBPX_ORYSA STANDARD; PRT; 429 AA.
 ID CBPX_ORYSA AC P52712;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Serine carboxypeptidase-like precursor (EC 3.4.16.-).
 GN CBP31.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Yukihikari;
 RA Washio K., Ishikawa K.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 CC -! CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 CC broad specificity.
 CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D17587; BAA04511.1; -
 CC PIR; T03607; T03607.
 CC DR HSSP; P00729; IYSC.
 CC DR MEROPS; S10.009; -
 CC DR Gramene; P52712; -
 CC DR InterPro; IPR000379; Ser esters site.
 CC DR InterPro; IPR001563; Serine carboxpept.
 CC DR Pfam; PF00450; serine carboxpept; 1.
 CC DR PRINTS; PR00724; CRBOXYPTASEC.
 CC DR ProDom; PD001189; Serine carboxpept; 1.
 CC DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 CC DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 CC KW Hydrolase; Carboxypeptidase; Glycoprotein; Signal.
 FT SIGNAL 1 ?
 FT CHAIN 1 ?
 FT ACT_SITE 148 148 SERINE CARBOXYPEPTIDASE-LIKE.
 FT ACT_SITE 336 336 BY SIMILARITY.
 FT ACT_SITE 393 393 BY SIMILARITY.
 FT BINDING 339 339 SUBSTRATE (BY SIMILARITY).
 FT CARBOHYD 76 76 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 429 AA; 47746 MW; 1D5A68544325BB1 CRC64;

Alignment Scores:
 Pred. No.: 6.28e-25 Length: 429
 Score: 473.00 Matches: 130
 Percent Similarity: 45.49% Conservative: 77
 Best Local Similarity: 28.57% Mismatches: 178

Query Match:	15.63%	Indels:	70
DB:	1	Gaps:	16
US-09-712-338-1 (1-1668) x CBPX_ORYSA (1-429)			
QY	208	TACTCTGGATAT-----GTGACACCTCTCCGAGTCCCATACCTTCTCTGTTCTTC	261
DB	17	HisAlaGlyTyrTyrArgLeuProAsnThrHisAspAlaArgLeuPheTyrPhePhe	36
QY	262	GAAGCCAGACATAACCA-----GAACTGCACCTATCATTTGTTGTTGAATGGGCGCT	318
DB	37	gluSerArgGlySerGlyGluAspAspProValValIleThrLeuThrGlyGlyPro	56
QY	319	GGAAGCGATTCTTTGATCGGCTCTTCGAAGAGTTGGCGCTTGCCTATGTCATTCGACT	378
DB	57	GlyCysSerSerGluLeuAlaLeuPheTyrGluAsnGlyProPheHisIleAlaAspAsn	76
QY	379	TTTGATGACTATACCAACCTCCTCGTGGAAACGAGGTCTCCATTTACTATTCTCTGCC	438
DB	77	MetSerLeuValTyrAsnAspPheGlyTyrAspGlnGluSerAsnLeuIleTyrValAsp	96
QY	439	CAGCCATTGGAGTCGGCTTTTCATATAGTATACGGTTGATGGTCCATTACCCCTGTA	498
DB	97	GlnProThrGlyThrGlyPheSerTyrSerSerAsnProArgAspThrArgHisAspGlu	116
QY	499	ACTGGGTGTCGAAATTCAGAGCTTTCAGAGAGTTTCAGGCGCGGTACCCCAACATTGAT	558
DB	117	AlaGly-----ValSerAsnAspLeuTyrAla-----	125
QY	559	GCCACTCTGATCGATPACTACCAATCTCCGAGAGCGGCTTGGAGATCTCGAAGGA	618
DB	126	-----PheLeuGlnAla 129	
QY	619	TTCTTAGTACTACCTAGTCTGGACTTAGGTGCAGTCTAAGACTTCAGTCTAGTCTAGG	678
DB	130	PheThrGluHisProAsnPhe-----AlaLysAsnAspPheTyrIleThr 145	
QY	679	ACGAGAGTACTGAGGCGCACATGCTCTGATCTTCATTCATTCATTTTACGAGCAGAT	738
DB	146	GlyGluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValTyrLysGlyAsn	165
QY	739	GAGAAATTCGCAACGGTAGTGTATATGTTGTTAGCTTAATTTCACTCTCTGGAATT	798
DB	166	Lys-----AsnSerGluGlyIleHisIleAsnLeuLysGlyPheAlaIle 180	
QY	799	ATTACGGCATCATCGACGAGCGATCCAGGCGCTTACTACCTGAATTCGCTGTGAC	858
DB	181	GlyAsnGlyLeuThrAspProAlaIleGlnTyrLysAlaTyrThrAspTyrSerLeuAsp	200
QY	859	AATACCTACGGTATCAAGCTGTCAACGAGACCGCTTCAACACTACATGAAGTTGCCAAC	918
DB	201	-----MetGlyLeuIleThrLysSerGlnPheAsnArgile-----Asn 213	
QY	919	CAATGCCAAATGTTGCCAGATTTGATTTTCACCTGCAACACAGCAACCGC-----ACC	975
DB	214	LysIleValProThrCysGluLeuAlaIleLysLeuCysGlyThrSerGlyThrIleSer	233
QY	976	GCATTAGCTGACTACGCTCTGCGCGGACGACCAACATGTCAGGACATGTTGAG	1035
DB	234	CysLeuGlyAlaTyrValValCys-----AsnLeuIleCysSerSerIleGlu 249	
QY	1036	GGCCCATACTACGCTTCTGCTGCTGTTGTATGATATTCGGCATCCA-----	1086
DB	250	-----ThrIleIleGlyLysLysAsnTyrTyrAspIleArgLysProCysValGly 266	
QY	1087	-----TATGATACCGCATCTCCGCAAGTTATTACAAATTTCTGGCAAGGAC 1137	
DB	267	SerLeuCysTyrAsp-----LeuSerAsnMetGluLysPheLeuGlnLeuLys 282	
QY	1138	TCTGTATGACCTATCGCGTC-----AACATCAACTACACCGACCTCCATATGACGTC	1194
DB	283	SerValArgGluSerLeuGlyValGlyAspIleGlnPheValSerCysSerProThrVal	302

RESULT 9

CBP3_ORYSA	STANDARD;	PRT;	500 AA.
ID	CBP3_ORYSA		
AC	P37891;		
DT	01-OCT-1994 (Rel. 30, Created)		
DT	01-OCT-1994 (Rel. 30, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Serine carboxypeptidase III precursor (EC 3.4.16.5).		
GN	CBP3.		
OS	Oryza sativa (Rice).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Eriactoidae; Oryzae; Oryza.		
OX	NCBI_TaxID=4530;		
RN	(1)		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Yukihikari; TISSUE=Seed;		
EX	MEDLINE=92329723; PubMed=1627776;		
RA	Washio K.; Ishikawa K.;		
RT	"Structure and expression during the germination of rice seeds of the		
RT	gene for a carboxypeptidase."		
RL	Plant Mol. Biol. 19:631-640(1992).		
CC	!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a		
CC	broad specificity.		
CC	!- SUBUNIT: Monomer (Probable).		
CC	!- INDUCTION: BY GIBBERELIC ACID (GA). INHIBITED BY ABSICISIC ACID		
CC	(ABA).		
CC	!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL; D10985; BAA01757.1; -		
DR	FIR; S22530; S22530.		
DR	HSSP; P00729; LYSC.		
DR	MEROPE; S10.009; -		
DR	Gramene; P37891; -		
DR	InterPro; IPR000379; Ser esters site.		
DR	InterPro; IPR001563; Serine carbpept.		
DR	Prints; PF00450; serine carbpept; 1.		
DR	PRINTS; PD00724; CRBOXYPTASEC.		
DR	ProDom; PD001189; Serine carbpept; 1.		
DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.		
DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.		

KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.

FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 73 BY SIMILARITY.
 FT CHAIN 74 484 SERINE CARBOXYPEPTIDASE III.
 FT PROPEP 485 500 BY SIMILARITY.
 FT ACT SITE 216 216 BY SIMILARITY.
 FT ACT SITE 404 404 BY SIMILARITY.
 FT ACT SITE 461 461 BY SIMILARITY.
 FT BINDING 407 407 SUBSTRATE (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 500 AA; 55446 MW; AE455E2780147DB8 CRC64;

Alignment Scores:

Pred. No.: 1.13e-23 Length: 500
 Score: 455.00 Matches: 143
 Percent Similarity: 43.16% Conservative: 84
 Best Local Similarity: 27.19% Mismatches: 217
 Query Match: 15.03% Indels: 82
 DB: 1 Gaps: 20

US-09-712-338-1 (1-1668) x CBP3_ORYSA (1-500)

QY 19 CTCTAGTCTACCTTGGTGTGTCAGCCAGTGGGCC-----CTTCCA----- 60
 DB 10 LeuLeuValValValLeuAlaAlaSerAlaCysAlaGluGlyLeuArgLeuProArgAsp 29
 QY 61 GGAAGTACACCGCGTCCGTCGTAGAGA-----CAGCTACCCAAAG 102
 DB 30 AlaLysPheProAlaAlaGlnAlaGluArgLeuLeuArgSerLeuAsnLeuLeuProLys 49
 QY 103 -----AACCCACCGGG-----GTCAAGACTTTTACAAACCGCAAAATGTCACC 147
 DB 50 GluAlaGlyProThrGlyAlaGlyAspValProSerValAlaProGlyGluLeuLeuGlu 69
 QY 148 ATCCGCTACAGGAACCGCGG---GCAGAGGGGCTCGACACTACCCCGGGTGTCAA 204
 DB 70 ArgArgValThrLeuProGlyLeuProGlnGlyValGlyAsp-----LeuGly 85
 QY 205 TCTACTCTGGATAT-----GTCGACACTCTCCGAGTCCCATACCTTCTCTGGTTC 258
 DB 86 HisHisAlaGlyTyrArgLeuProAsnThrHisAspAlaArgMetPheTyrPheLeu 105
 QY 259 TTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTGTGGTGAATGGTGGCCCT 318
 DB 106 PheGluSerArgGlyLysGlyAsp---ProValValIleTyrLeuThrGlyGlyPro 124
 QY 319 GGAAGCGATTCTTTCATCGTCTCTCGAAGAGTTGGCCCTTGCATGTCAATTCGACT 378
 DB 125 GlyCysSerSerGluLeuAlaValPheTyrGluAsnGlyProPheThrIleSerAsnAsn 144
 QY 379 TTTGATGACTACATCAACCCCTCACTCGTGGACGAGGTCTCCAAATTACTATTCTGTCC 438
 DB 145 MetSerLeuAlaTyrAsnLysPheGlyTyrAspThrIleSerAsnIlePheValAsp 164
 QY 439 CAGCCATTGGAGTCGGCTTTTCATATGATGATACGTTGATGGTTCATTAACCCCTGTA 498
 DB 165 GlnProThrGlyThrGlyPheSerTyrSerSerAspAspArgAspThrArgHisAspGlu 184
 QY 499 ACTGGGTCGTGCGAAATTCGAGCTTTGAGAGTTTCAGGAGTTTCAGGCGCGTACCACTTGTAT 558
 DB 185 ThrGlyVal----- 187
 QY 559 GCCACTCTGATCGATACCAATCTTCGCGAGAGCGGCTTGGAGATCCTGCAAGGA 618
 DB 188 -----SerAsnAspLeuTyrSerPheLeuGlnVal 197
 QY 619 TTCCTTAGTGACTACTAGCTTGGACTTAGGGTGCAGTCTAAGGACTTCAGTCTATGG 678
 DB 198 PhePheLysLysHisProGluPhe-----AlaLysAsnAspPhePheIleThr 213
 QY 679 ACGGAGACTATGGGGGCACTATGGTCTGCTCATCTTCAATCAATTTTACGAGCAAT 738
 DB 214 GlyGluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValHisGlnGlyAsn 233

QY 739 GAGAGAAATGCCAACCGTAGTGTATGTTGTTTCAGCTTAATTTCAACTCTCTGGGAATT 798
 DB 234 Lys-----AlaAsn-----GluGlyIleHisIleAsnLeuLysGlyPheAlaIle 248
 QY 799 AATAACGGCATCATGACGAGGCGATCCAGGCCCTTACTACCTGAATTCGTTGTAAC 858
 DB 249 GlyAsnGlyLeuThrAspProAlaIleGlnTyrLysAlaTyrThrAspTyrAlaLeuAsp 268
 QY 859 AATACCTACGGTATCAAGGCTGTCAACGAGACCGTCTACAACTTACATGAATTTGCCAAC 918
 DB 269 -----MetAsnLeuLysLysSerAspTyrAspArgile-----Asn 281
 QY 919 CAAATGCCAAATGGTTCGAGGATTTGATTTCCACTGCAACAGACAAACCGCACCGCA 978
 DB 282 LysPheIleProProCysGluPheAlaIleLysLeuCysGlyThrAsnGlyLys----- 299
 QY 979 TTAGCTGACTACGCTCTGCGCGAAGCCACCAACTGTGAGGACACATGTTGAGGGG 1038
 DB 300 -----AlaSerCysMetAlaAlaTyrMetValCysAsnSerIlePheSerSer 315
 QY 1039 CCATACTACGCTTCTGCTGTGTGTATGATTTTCGGCATCCATATGATGATCCCG 1098
 DB 316 IleMetLysLeuValGlyThrLysAsnTyrTyrAspValArgLysGluCysGluGlyLys 335
 QY 1099 ACTCCGCCCAAGTTATTACAA-----AAATTTCTGGCAAAGACTCTGTCTATGGACGT 1152
 DB 336 LeuCysTyrAspPheSerAsnLeuGluLysPheGlyAspLysAlaValLysGluAla 355
 QY 1153 ATCGGCGCT-----AACATCAACTACACCCAGTCCCAATATGACGTCTACTACGTTCCAG 1209
 DB 356 IleGlyValGlyAspLeuGluPheValSerCysSerThrThrValTyrGlnAlaMetLeu 375
 QY 1210 CAAACAGGCGACTTGTCTGCGCCCACTTCATCGAA-----GACCTCGAGGAGATCCTTGTCT 1266
 DB 376 ThrAsp-----TyrMetArgAsnLeuGluValGlyIleProAlaLeuLeuGlu 391
 QY 1267 CTCCTCGGTCGTCTCTCTCATCTATGCGAGCGCGATTAATCTGCAACTGGTTCCGC 1326
 DB 392 AspGlyIleAsnValLeuIleTyrAlaGlyGluTyrAspLeuIleCysAsnTrpLeuGly 411
 QY 1327 GGTACGCGCGTTTCCCTCGCTGCGAACTACTCCCAAGCGCCCGCTTCGAGGCGAGG 1386
 DB 412 AsnSerArgTyrValHisSerMetGluTrpSerGlyGlnLysAspPheValSerSerHis 431
 QY 1387 TACAGCCCTCTGAAAGTCAACGCGCTCGAGTATGGGAAACTCGCGAGTATCGTAATTC 1446
 DB 432 GluSerProPheValValAspGlyAlaGluAlaGlyValLeuLysSerHisGlyProLeu 451
 QY 1447 TCCTTCACTCGGTCATGACGAGCGCATGAGTCCCATCTACTACGAGCCCATCCGCTCC 1506
 DB 452 SerPheLeuLysValHisAsnAlaGlyHisMetValProMetAspGlnProLysAlaSer 471
 QY 1507 CTGCAATTTGTTAAACCGG 1524
 DB 472 LeuGluMetLeuArgArg 477
 RESULT 10
 CBP3_HORVU
 ID CBP3 HORVU STANDARD; PRT; 508 AA.
 AC P21529;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-MIII).
 GN CBP3 OR CXP.3;
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
 OC Triticeae; Hordeum.
 NC NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Himalaya; TISSUE=Aleurone;
RA Rocher A., Lok P., Cameron-Mills V., von Wettstein D.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 81-491.
RX STRAIN=cv. Gula;
RY MEDLINE=90315015; PubMed=2639682;
RA Soerensen S.B., Svendsen I., Bredam K.;
RT "Primary structure of carboxypeptidase III from malted barley.";
RL Carlsberg Res. Commun. 54:193-203(1989).
CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -!- ENZYME REGULATION: INHIBITED BY MERCURIC IONS.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOSPERM.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED MAINLY IN THE ALEURONE AND, TO A
CC LESSER EXTENT IN THE EMBRYO, THROUGHOUT THE 5-DAYS GERMINATION
CC PERIOD EXCLUSIVELY, WITH A MAXIMAL LEVEL AT 3 DAYS. ALSO FOUND IN
CC THE ROOTS AND SHOOTS OF THE GROWING SEEDLING.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y09604; CAA70817.1; -
DR HSSP; P00729; 1CPY.
DR MEROPS; S10.009; -
DR InterPro; IPR000379; Ser esters site.
DR InterPro; IPR001563; Serine_carbpept.
DR Pfam; PF04507; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine carbpept; 1.
DR PROSITE; PS001131; CARBOXYPEPT SER SER; 1.
DR PROSITE; PS00560; CARBOXYPEPT SER HIS; 1.
KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 19
FT PROPEP 20 80
FT CHAIN 81 491
FT PROPEP 492 508
FT MOD RES 81 81
FT ACT_SITE 223 223
FT ACT_SITE 411 411
FT ACT_SITE 468 468
FT BINDING 414 414
FT CARBOHYD 151 151
FT VARIANT 265 265
SQ SEQUENCE 508 AA; 56362 MW; 70C6751D78D40AB6 CRC64;

Alignment Scores:
Pred. No.: 1,96e-22 Length: 508
Score: 437.00 Matches: 134
Percent Similarity: 41.98% Conservative: 78
Best Local Similarity: 26.53% Mismatches: 225
Query Match: 14.44% Indels: 68
DB: 1 Gaps: 15

US-09-712-338-1 (1-1668) x CBP3_HORVU (1-508)
QY 34 TTGGTTCAGCAGTTGGGCGCTTCAGGAAGTACACCGCGGCCCTCGGTAGAACACAG 93
Db 40 LeuileArgAlaLeuAsnLeuLeuProLysAspSerSerSerSerGlyArgHisGly 59
QY 94 CTACCAAGAACCCCGGGGTCAAGACTTTTACAAACCGCAACAAATGTCACCATCCGG 153
Db 60 ---AlaArgValGlyGluGlyAsnGluAspValAlaProGlyGlnLeuLeuGluArg 78
QY 154 TACAAGAACCCCGG---GCAGAGGCGTCTCGAGACTACCCCGGTGTCAAATCCTAC 210

Db 79 ValThrLeuProGlyLeuProGluGlyValAlaAsp-----LeuGlyHisHis 94
QY 211 TCTGGATAT-----GTCGACACCTTCCGAGTCCCATACCTTCTTCTGTTCTTCAA 264
Db 95 AlaGlyTyrTyrArgLeuProAsnThrHisAspAlaArgMetPheTyrPhePheGlu 114
QY 265 GCAGACATAAACCCAGAAATGACCATATCACATTGTGTGGTGAATGATGCCCTCGGAAGC 324
Db 115 SerArgGlyLysLysGluAsp---ProValValIleThrLeuThrGlyGlyProGlyCys 133
QY 325 GATTCTTTGATCGTCTCTCGAAGAGTTCGGCCCTTCGCCATGTCATTCGACTTTTGCAT 384
Db 134 SerSerGluLeuAlaValPheTyrGluAsnGlyProPheThrIleAlaAsnAsnMetSer 153
QY 385 GACTTACATCAACCTCTCGTGACGAGGTCTTCCAAATTCTTCTTCTTCTTCTTCCAGCCA 444
Db 154 LeuValIleAsnLysPheGlyTyrAspLysIleSerAsnIleIlePheValAspGlnPro 173
QY 445 TTGGAGTCGGCTTTTCATATAGTATGATACGGTTGATGGTCCATTACCCCTGTAACCTGG 504
Db 174 ThrGlyThrGlyPheSerTyrSerSerAspAspArgAspThrArgHisAspGluThrGly 193
QY 505 GTCGTCGAAAATTCGAGCTTTGCAGAGATTTCAGGGCCGGTACCCCAACCATTTGATGCCACT 564
Db 194 Val-----
QY 565 CTGATCGATACTACCAATCTTCCGCGAGAGCCGCTTGGGAGATCTCTCAGAGGATTCCTT 624
Db 195 -----SerAsnAspLeuTyrAspPheLeuGlnValPhePhe 206
QY 625 AGTGGACTACCTAGCTTCGAGCTTAGGGTGCAGTCTAAGGACTTCAGTCTTAGGACGCGAG 684
Db 207 LysLysHisProGluPhe-----IleLysAsnAspPhePheIleThrGlyGlu 222
QY 685 AGCTATGGAGGCGACTATGCTGCTCATTTCTTCAATCTTTTACGACGAGATGAGAGA 744
Db 223 SerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValHisGlnGlyAsnLysLys 242
QY 745 ATTGCAACGGTAGTGTTAATGGTGTTCAGCTTAATTTCACTCTCTGGGATTTATTAC 804
Db 243 -----AsnGluGlyThrHisIleAsnLeuLysGlyPheAlaIleGlyAsn 257
QY 805 GGCATCATCGAGGCGGCTCCAGCCCTTACTACCTGCTGAGTTCGTTGACCAATACC 864
Db 258 GlyLeuThrAspProAlaIleGlnTyrLysAlaTyrThrAspTyrAlaLeuGluMetAsn 277
QY 865 TACGGTATCAAGGCTGTCAACGAGACCGCTCTACAACTACATGAGTTTGCACCAACAAATG 924
Db 278 LeuIleGlnLysAlaAspTyrGluArgile-----AsnLysPhe 290
QY 925 CCAATGGTTCGACGAGTTTGTATTTCCACTCCACTCCAAACAGACAAACCGC---ACCGCATTA 981
Db 291 IleProProCysGluPheAlaIleLysLeuCysGlyThrAsnGlyLysAlaSerCysMet 310
QY 982 GCTGACTAGCCCTCTGCGCCGAGCCCAACACATGTGCAGGAGCAATCTGAGGGGCCA 1041
Db 311 AlaAlaTyrMetValCysAsnThrIlePheAsnSerIleMetLysLeuValGlyThrLys 330
QY 1042 TACTACGCGCTTTGCTGCTGCTGCTGATGATTTCCGCATTCATATGATGATGATGATGAT 1101
Db 331 AsnTyr-----TyrAspValArgLysGluCysGluGlyLysLeu 343
QY 1102 CGCCCAAGTTATTACAC-----AAATTTCTGGCAAGAGTCTGTCTGATGACGCTATC 1155
Db 344 CysTyrAspPheSerAsnLeuGluLysPhePheGlyAspLysAlaValArgGlnAlaIle 363
QY 1156 GGCGTC---AACATCAACTACCCAGTCCAAATGATGATGATGATGATGATGATGATGATGAT 1212
Db 364 GlyValGlyAspIleGluPheValSerCysSerThrSerValTyrGlnAlaMetLeuThr 383
QY 1213 ACAGGCGACTTGTCTGGGCCCACTTCATCGAA---GACCTCGAGAGATCTTCTGCTCTC 1269
Db 384 Asp-----TrpMetArgAsnLeuGluValGlyIleProAlaLeuLeuGluAsp 399

FT	CARBOHYD	307	307	N-UNLINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	346	346	N-UNLINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	25	25	R -> H (IN REF. 2).
FT	CONFLICT	284	284	L -> F (IN REF. 3 AND 4).
FT	CONFLICT	287	287	F -> L (IN REF. 3).
FT	CONFLICT	398	398	H -> R (IN REF. 3 AND 4).
FT	CONFLICT	422	422	F -> L (IN REF. 2).
FT	CONFLICT	435	435	A -> V (IN REF. 3 AND 4).
FT	CONFLICT	438	438	F -> S (IN REF. 2).
SQ	SEQUENCE	476 AA;	54110 MW;	2D966683A4F3FD01 CRC64;

Alignment Scores:	
Pred. No.:	3,45e-17
Score:	360.50
Percent Similarity:	39.22%
Best Local Similarity:	26.21%
Query Match:	11.91%
DB:	1
DB:	23

US-09-712-338-1 (1-1668) x CPVL_HUMAN (1-476)			
QY	50	GGGCCCTTCCAGGAAGTACACCGGCTCCGTGGTAGAGACAGCTACCCAGAACCCTCA	109
DB	22	GlyLeuPheArgSerLeuTyrArgSerValSerMet	Pro 34
QY	110	CGGGGTCAAGACTCTTACAACCGCAACAATGTCACCA-	160
DB	35	ProLysGlyAspSerGlyGlnProLeuPheLeuThrProTyrIleGluAlaGlyLysIle	54
QY	161	AACCCGGGCGCAGGGCGTCCGAGACTACCCCG-	210
DB	55	GlnLysGlyArgGluLeuSerLeuValGlyProPheProGlyLeuAsnMetLysSerTyr	74
QY	211	TCTGGATAT-----GTCGACACCTCTCCGAGTCCCATCTCTCTCGGTCTCTCGAA	264
DB	75	AlaGlyPheLeuThrValAsnLysThrTyrAsnSerAsnLeuPheThrPhePro	94
QY	265	GCCAGACATAACCCAGAAATCGACCTATCATTTGGTTGAATGGTGGCCCTCGAAGC	324
DB	95	AlaGlnIleGlnProGluAspAlaProValLeuLeuTyrLeuGlnGlyProGlyGly	114
QY	325	GATTCTTTGATCGGCTCTTCGAACAGTGGGCCCTTCCCATGTCAATTCGACTTTGAT	384
DB	115	SerSerMetPheGlyLeuPheValGluHisGlyProTyrValValThrSerAsnMetThr	134
QY	385	GACTACATCAACCCCTCACTCTGGAAACGAGGTCTCCAATTTACTATTCTGTGCCAGCA	444
DB	135	LeuArgAspArgAspPheProIleThrThrLeuSerMetLeuTyrIleAspAsnPro	154
QY	445	TTGGAGTCGGCTTTTCATATAGTAGTACGGTTGATGGTCCATTAACCTGTAACTGGG	504
DB	155	ValGlyThrGlyPheSerPheThrAspAspThrHisGly-	167
QY	505	GTCCTGAAAAATTCGAGCTTTTCGAGGAGTTCAGGGCCGGTACCAACCATTTGTCGCACT	564
DB	168	-----Tyr	168
QY	565	CTGATCGATACTACCAATCTTCCGACAGGCGCTGGGAGATCTCTCAAGGATTCCTT	624
DB	169	AlaValAsnGluAspAspValAlaArgAspLeuTyrSerAlaLeuIleGlnPhePhe	187
QY	625	AGTGGACTACCTAGCTTGGACTTAGGGTCAGTCTAAGGACTTCAGTCTATGACCGGAG	684
DB	188	-----GlnIlePheProGluTyrLysAsnAspPheTyrValThrGlyGlu	203
QY	685	AGCTATGGAGGCACTATGTCTCTGCAATCTTCAATCATTTTACGAGCAGACAGTACAGA	744
DB	204	SerTyrAlaGlyLysTyrValProAlaIleAlaHisLeuIleHisSerLeuAsn	221
QY	745	ATTGCCAACGGTAGTGTAAATGGTTCAGCTTAATTTCAACTCTCTCGGAATATTATTAAC	804
DB	222	-----ProValArgGluValLysIleAsnLeuAsnGlyIleAlaIleGlyAsp	237

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U28730; AAA68259.1; -
 DR PIR; T16606; T16606.
 DR HSSP; P10619; 11VY.
 DR MEROPS; S10.002; -CRO2009.
 DR WormPep; K1082.2; CRO2009.
 DR InterPro; IPR000379; Ser_estr_ site.
 DR Pfam; IPR001563; Serine carbpept.
 DR Pfam; PF00450; serine carbpept; 1.
 DR PRINTS; PR00724; CRBOXVPTASEC.
 DR ProDom; PD001189; Serine carbpept; 2.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 DR KW Hypothetical protein; Hydrolase; Carboxypeptidase; Glycoprotein;
 KW Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 470
 FT ACT_SITE 169 169
 FT ACT_SITE 380 380
 FT ACT_SITE 441 441
 FT ACT_SITE 132 132
 FT CARBOHYD 132 132
 FT CARBOHYD 316 316
 FT CARBOHYD 396 396
 FT CARBOHYD 470 AA; 53158 MW; CCC2DACB75EF30FC CRC64;
 SQ

Alignment Scores:
 Pred. No.: 7.58e-17 Length: 470
 Score: 355.50 Matches: 140
 Percent Similarity: 37.59% Conservative: 69
 Best Local Similarity: 25.18% Mismatches: 194
 Query Match: 11.74% Indels: 153
 DB: 1 Gaps: 22

US-09-712-338-1 (1-1668) x YSS2_CAEEL (1-470)
 QY 13 GAATTTCTCAGTGTACCTTGGTTCAGCCAGTTGGGCCCTTCCAGGAGTACACCG 72
 Db 2 LysLeuLeuSerIleLeuPheIleValSerTyrSerPheCysLeuAlaAlaPro 21
 QY 73 GCCTCGCTCGGTAGAAGACAGCTACCCAGAACCCACCACCGGGGTCAAGACTCTTACAAC 132
 Db 22 AlaThrAspLysValAsnAspLeuPro----- 30
 QY 133 GCAAAACATGTCCACATCCGGTCAAGGAACCCGGGCGAGAGGGCTCTCGAGACTAC 192
 Db 31 -----GlyLeuThrPhe-----Thr 35
 QY 193 CGGGTGTCAATCCTACTCTGGATATGTCGACACCTCTCCGAGTCCCATACCTTCTC 252
 Db 36 ProAspPhePheHisTyrSerGlyTyrLeuArgAlaThrPheAspLysTyrLeuHisTyr 55
 QY 253 TGGTCTCTTCAAGCCAGACATACCCAGAACTGCACCTATCATCATTGGTTGAATGGT 312
 Db 56 TrpLeuThrGluSerSerArgAlaProThrGlnAspProLeuValLeuThrLeuAsnGly 75
 QY 313 GGCCCTCGAGCCGATTCTTTGATCGGCTCTTCCAGAGTTGGCCCTTGCATGTCAAT 372
 Db 76 GlyProGlyCysSerSerLeuAspGlyLeuIleGluLeuGlyProPheHisValys 95
 QY 373 TCGACTTTTGATGAC---TACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTACTA 429
 Db 96 AspPheGlyAsnSerIleTyrTyrAsnGluTyrAlaTrpAsnLysPheAlaAsnValLeu 115

QY 430 TTCTGTCCAGCAGCATTTGGAGTCGGCTTTTTCATATAGTATGATACGGTTGATGGTCCATT 489
 Db 116 PheLeuGluSerProAlaGlyValGlyTyrSerTyrSerThrAsnLeuThrVal 135
 QY 490 AACCTCTTAATCGGGTCTGTCGAAAATTCGAGGCTTTTCAGGAGTTCAGGCGCGGTACCCA 549
 Db 136 SerAspAspGluValSerLeuHisAsn----- 144
 QY 550 ACCATTGATGCCACTCTGATCGTACTACCAATCTTCCGAGAGCGGCTTGGGATC 609
 Db 145 -----TyrMetAla 147
 QY 610 CTGCAAGGATTCCTTAGTGGACTACTAGCTTGGACTCTAGGTCGAGTCTTAAGGACTTC 669
 Db 148 LeuLeuAspPheLeuSerLysPheProGluTyrLysGlyArg-----AspPhe 163
 QY 670 AGTCTATGGACGAGAGCTATGGAGGCACTATGGTCTCGCATCTTCTCAATCAITTTTAC 729
 Db 164 TrpIleThrGlyGluSerTyrAlaGlyValTyrIleProThrLeu----- 178
 QY 730 GAGCAGAAATGAGAGAAATTCGCAACGGTAGTGTATGTTGCTTACGCTTAATTTCAACTCT 789
 Db 179 -----AlaValArgIleLeuAsnAspLysLysAsnPro-----AsnPhelysGly 194
 QY 790 CTGGGAATTATTACGGCATCATC-----GACGAGCGGATCCAGGCCCTTAC 837
 Db 195 ValAlaIleGlyAsnGlyAlaLeuAsnPheProAsnAsnTyrAsnThrMetValProPhe 214
 QY 838 TACCCTGAAATCGCTGTGAACAATACCTACGGTATCAAGCGTGTCAAGAGACCGCTTAC 897
 Db 215 Tyr-----TyrTyrHisAlaLeuValArgAspLeuTyr 226
 QY 898 AACTACATGAAGTTGCCAACCAATGCCAATGCTGCGAGATTTGATTTCCACTCG 957
 Db 227 Asn-----AspIleAlaArgAsnCysCysAsnAsnIleGlyThrCys 241
 QY 958 AAACAGACAAACCGCACCGCATTAGCTGACTACGCCCTCTGGCGCCGAGCCACCAACATG 1017
 Db 242 Asp-----IleTyrSerLysPhePheAspProAsn----- 251
 QY 1018 TGCAGGAGCAATGTTGAGGGGCCATACCTACGCTTTGCTGCTGGTGTGTATGATATT 1077
 Db 252 CysArgAspLysVal-----IleAsnAlaLeuAspGlyThrAsnGluLeuAsnMet 268
 QY 1078 CGGATCCTATATGATGATACCGGCTCGCCCAAGTTATTACAAC----- 1119
 Db 269 TyrAsnLeuTyrAsp-----ValCysTyrTyrAsnProThrThrAsnLeuLys 284
 QY 1120 AAATTTCTGGCAAAGGACTCTGTCTGAGCGCTATCGCGCTC-----AAC 1164
 Db 285 LysAlaPheIleGluArgGlnMetArgIleAlaValGlyLeuProAlaArgLysHisAsn 304
 QY 1165 ATCAACTACACC-----CAGTCCAAATATGACGCTCTACTAC----- 1200
 Db 305 AlaAlaThrThrValProLeuCysAlaGlnThrAsnAsnThrHisValTyrLeuAsnArg 324
 QY 1201 -----GCTTTCAGCAACACAGGC 1218
 Db 325 AlaAspValArgLysSerLeuHisIleProSerSerLeuProAlaTrpGluGluCysSer 344
 QY 1219 GACTTTGCTGGCCCAACTTC-----ATCGAAGACCTCGAGGAG 1257
 Db 345 AspGlnValGlyLysAsnValValThrHisPheAsnValIleProGluPheGlnThr 364
 QY 1258 ATCCTGTCTCCCGGTGCTGCTCCCTCATCTATGCGCAGCCGCTTACATCTGCAAC 1317
 Db 365 MetIleAlaGlyIleIleLeuValTyrAsnGlyAspValAspThrAlaCysAsn 384
 QY 1318 TGGTTCGGCGGTGAGCGGCTTCCCTCGCTGCAACTACTCCCAAGCGCGCCAG----- 1371
 Db 385 SerIleMetAsnGlnGlnPheLeuThrSerLeuAsnLeuThrValLeuGlyGluGlnGlu 404
 QY 1372 -----TTCCGAAGCGCAGGGGTACAGGCCCTCGAAAGTCAACGCGGTCTCAG 1416

Db 405 LysValAsnGluAlaTrpHisTyrSerGlyGlnThrGlyThrAlaValAlaGlyPhe--- 423
 QY 1417 TATGGGAACTCGAGTARGTAATTCCTTCCTACTCGCTATGAGGAGGCAT 1476
 Db 424 -----GlnThrLysPheAlaGlyAsnValAspPheLeuThrValArgGlySerGlyHis 441
 QY 1477 GAGTCCCACTACAGCCCATCGCTCCCTGCA---TTGTTTAAAC 1521
 Db 442 PheValProGluAspLysProLysGluSerGlnGlnMetIlePheAsn 457

RESULT 14
 CP22_HORVU
 ID_CP22_HORVU STANDARD; PRT; 436 AA.
 AC P55748;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Serine carboxypeptidase II-2 precursor (EC 3.4.16.6) (CP-MII.2)
 DE (Fragment).
 GN EXP;2-2.
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=cv. Alexis; TISSUE=Grain;
 RX MEDLINE=94336715; PubMed=7520177;
 RA Dal Degán F., Rocher A., Caceron-Mills V., von Wettstein D.;
 RT "The expression of serine carboxypeptidases during maturation and
 RL germination of the barley grain.";
 RC Proc. Natl. Acad. Sci. U.S.A. 91:8209-8213(1994).
 CC -!- CATALYTIC ACTIVITY: Preferential release of a C-terminal arginine
 CC or lysine residue.
 CC -!- SUBUNIT: CARBOXYPEPTIDASE II IS A DIMER, WHERE EACH MONOMER IS
 CC COMPOSED OF TWO CHAINS LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE GERMINATING EMBRYO. LOW
 CC LEVELS IN THE DEVELOPING ALEURONE AND EMBRYO. ALSO FOUND IN THE
 CC ROOTS AND SHOOTS OF THE GROWING SEEDLING.
 CC -!- PTM: THE LINKER PEPTIDE IS ENDOPEPTIDOLYTICALLY EXCISED DURING
 CC ENZYME MATURATION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 or send an email to license@isb-sib.ch).

 EMEL; X78878; CAB59202.1; --
 DR HSP; P08819; 1WHT.
 DR InterPro; IPR000379; Ser estrs site.
 DR InterPro; IPR001563; Serine carboxpept.
 DR Pfam; PF00450; serine carboxpept; 1.
 DR PRINTS; PR00724; CRBOXYPTASEC.
 DR ProDom; PD001189; Serine carboxpept; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Multigene family.
 FT NON_TER 1
 FT CHAIN <1 256 SERINE CARBOXYPEPTIDASE II-2, CHAIN A.
 FT PROPEP 257 270 LINKER PEPTIDE (BY SIMILARITY).
 FT CHAIN 271 436 SERINE CARBOXYPEPTIDASE II-2, CHAIN B.
 FT ACT_SITE 149 149 BY SIMILARITY.
 FT ACT_SITE 350 350 BY SIMILARITY.
 FT ACT_SITE 403 403 BY SIMILARITY.
 FT DISULFID 56 313 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 253 281 BY SIMILARITY.

FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 436 AA; 48952 MW; E0F82D97E0C34DC9 CRC64;
 Alignment Scores:
 Pred. No.: 1.02e-16 Length: 436
 Score: 353.50 Matches: 123
 Percent Similarity: 41.54% Conservative: 76
 Best Local Similarity: 25.68% Mismatches: 173
 Query Match: 11.68% Indels: 107
 DB: 1 Gaps: 23

US-09-712-338-1 (1-1668) x CP22_HORVU (1-436)

QY 208 TACTCTGATATGTGACACCTCTCCGAG-----TCCCATACCTTCTTCTGCTTCTTC 261
 Db 16 TyrAlaGlyTyrValThrValSerGluAspArgGlyAlaAlaLeuPheTyrTrpPhePhe 35
 QY 262 GAAGCCAGACATAACCCAGAACTGCACCTATCACATTGTGGTTGAATGTGGCCCTGGA 321
 Db 36 GluAlaAlaHisAspProAlaSerLysProLeuLeuLeuTrpLeuAsnGlyGlyProGly 55
 QY 322 AGCGATTCTTTG---ATCGGTCTCTTCGAAGAGTTGGGCCCTTGGCCATGTCATTCGACT 378
 Db 56 CysSerSerIleAlaPheGlyValGlyGluGluValGlyProPheHisValAsnAlaAsp 75
 QY 379 TTGTATGAC---TACATCAACCTCTACTCTGTGACAGAGTCTCCAAATTTACTATTCTCTG 435
 Db 76 GlyLysGlyValHisMetAsnProTyrSerTyrAsnGlnValAlaAlaAsnIleLeuPheLeu 95
 QY 436 TCCAGCCATTGGGAGTGGCTTTTCATATAGTATGATGAGTTGATGGTTCATTAAACCT 495
 Db 96 AspSerProValGlyValGlyTyrSerTyrSerAsnThr----- 108
 QY 496 GTAACCTGGGTGCTCGAAAATTCGAGCTTTCGAGGAGTTTCAGGCGCGGTACCCACCAATT 555
 Db 109 SerAlaAspIleLeuSerAsn-----GlyAspGluArgThrAla 121
 QY 556 GATGCCACTCTGATGTACTACTCAATCTTCGCTTCTTCAATCAATTTTACGAGCAG 615
 Db 122 LysAspSerLeuValPheLeuThrLys-----Trp-----LeuGlu 133
 QY 616 GGATTCCTTAGTGAGTACTACCTAGCTTGGACTTAGGTGCAGTCTAAGAGCTTCAGTCTA 675
 Db 134 ArgPhe-----ProGlnTyrLysGluArgGluPheTyrLeu 145
 QY 676 TGCACGGAGAGTATGGAGGCACTATGCTGCTGCTTCTTCAATCAATTTTACGAGCAG 735
 Db 146 ThrGlyGluSerTyrAlaGlyHisTyrValProGlnLeuAlaGlnAlaIleLysArgHis 165
 QY 736 AATGAGAGAATTCGCAACGGTAGTGTAAATGGTGTGCTAGCTTAAATTCACCTCTCTGGA 795
 Db 166 HisGluAlaThrGlyAspLysSerIle-----AsnLeuLysGlyTyrMet 180
 QY 796 ATTATTAAACGGCATCATCGACGAGGCGATCCAGGCGCCCTTACTACCTGAATTCCTGTG 855
 Db 181 ValGlyAsnAlaLeuThrAspAspPhe-----HisAspHisTyrGlyIlePheGlnTyr 198
 QY 856 AACAAATACCTACGGTATCAACGGCTGTCAACGAGACCGCTACAACTACATGAGTTTCC 915
 Db 199 MetTrpThrThrGlyLeu-----IleSerAspGlnThrTyrLysLeuLeuAsnIlePhe 216
 QY 916 AACCAATGCCAAATGGTTC-----CAGGATTGATTTCACCTGCAACACAGACAAC 969
 Db 217 -----CysAspPheGluSerPheValHisThrSerProGlnCysAsp 230
 QY 970 CGCACCATTAGCTGATAGCTAGCCCTCTGCCCGAAGCCCAACATGTGCGAGGACAAT 1029
 Db 231 Lys-----IleLeuAspIleAla---SerThrGluAlaGlyAsnIleAspSerTyrSer 247
 QY 1030 GTTGAGGGGCA-----TACTAGCCTTTCCTGGT----- 1059
 Db 248 IlePheThrProThrCysHisSerSerPheAlaSerSerArgAsnLysValLysArg 267

QY 1060 ---CGTGGTGTATGATATTGGCATCCATATGATGACCGGACTCGCCAAAGT----- 1110
Db 268 LeuArgSerValGlyLysMeiGlyGluTyrAspProCysThrGluLysHisSerIle 287
QY 1111 ---TATTACAAATAATTCTGGCAAGAGACTCT-----GTCTGGAGCGTATCGGC--- 1158
Db 288 ValTyrPheAsnLeuHisGluValGlnLysAlaLeuHisValAsnProValIleGlyLys 307
QY 1159 -----GTCAACATCACTACACCCAGTCCCAATATGAC 1191
Db 308 SerLysTrpGluThrCysSerGluValIleAsnThrAsnTrpLysAspCysGluArgSer 327
QY 1192 GTCTACTACGCTTTCCAGCAAAAGCGGACTTGTCTGGCCCAACTTCATCGAAGACCTC 1251
Db 328 ValLeuHisIleTyr----- 332
QY 1252 GAGGAGATCCTTGCTCTCCCGTGGTCTCCCTCATCTATGCGGAGCGGATATCATC 1311
Db 333 HisGluLeuIleGlnTyrGlyLeuArgIleTrpMetPheSerGlyAspThrAspAlaVal 352
QY 1312 TGCAACTGGTTCGGCGGTTCAGCCGCTTCCCTCGCTCGCAACTACTCCCAAGCCGCCAG 1371
Db 353 -----IleProValThrSerThrArgTyrSerIleAspAla 364
QY 1372 TTCGAAGCGGAGGTACACGCCCTCGAAAGTC-----AACGGCGTCCGAGTAT 1419
Db 365 LeuLysLeuProThrValThrProTrpHisAlaTrpTyrAspAspGlyGluValGly 384
QY 1420 GGGGAAATCGCGAGTATGTAATTTCTTCCTTCATCTCGGCTATGAGGAGCGGATGAA 1479
Db 385 GlyTrpThrGlnGlyTyrLysGlyLeuAsnPheValThrValArgGlyAlaGlyHisGlu 404
QY 1480 GTCCATACTACACGCCATCCCTCCCTCGCAATGTTTAAACGGGACTATCTCGGT 1536
Db 405 ValProLeuHisArgProLysGlnAlaLeuThrLeuIleLysSerPheLeuAlaGly 423

RESULT 15
NF31_NAEFO STANDARD; PRT; 482 AA.
AC P42661;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Virulence-related protein NF314 (EC 3.4.16.-).
OS Naegleria fowleri
OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.
OX NCBI_TaxID=5763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEE;
RX MEDLINE=92267659; PubMed=1587609;
RT Hu W.-N., Kopachik W., Band R.N.;
RT "Cloning and characterization of transcripts showing
RT virulence-related gene expression in Naegleria fowleri";
RL Infect. Immun. 60:2418-2424(1992).
CC -!- FUNCTION: MAY BE REQUIRED BUT IS NOT SUFFICIENT FOR INCREASED
CC VIRULENCE.
CC -!- INDUCTION: BY GROWTH ON MAMMALIAN CELLS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).
CC
CC EMBL; M88397; AAA29384.1; --
DR PIR; A43828; A43828.
DR HSSP; P10619; 11VY.
DR MEROPS; S10.UFW; --
DR InterPro; IPR000379; Ser_estrs_site.

InterPro; IPR001563; Serine_carbpept.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carbpept; 2.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
KW Hydrolase; carboxypeptidase. BY SIMILARITY.
FT ACT SITE 163 163
FT ACT SITE 399 399 BY SIMILARITY.
FT ACT SITE 459 459 BY SIMILARITY.
SQ SEQUENCE 482 AA; 53848 MW; 0B83049C1A8A0908 CRC64;

Alignment Scores:
Pred. No.: 1.56e-16 Length: 482
Score: 351.00 Matches: 139
Percent Similarity: 35.19% Conservative: 57
Best Local Similarity: 24.96% Mismatches: 165
Query Match: 11.60% Indels: 196
DB: 23 Gaps: 23

US-09-712-338-1 (1-1668) x NF31_NAEFO (1-482)
QY 163 CCCGGGCGAGAGGGCGTCTGCGAGACTACCCCGGGTGTCAATCTACTCTGATATGTC 222
Db 20 ProGlyLeuSerGly-----AsnIleGlyValLysSerTyrThrGlyTyrLeu 35
QY 223 -----GACACTCTCCCGAGTCCCATCTCTTCGTCTTCTCGAAGCCAGACATAAC 276
Db 36 LeuAlaAsnAlaThrArgGlyArgTyrLeuPheTyrTrpPheGluSerMetArgAsn 55
QY 277 CCAGAAATCGCACCTATCACATTGGTGTGAATGTGGCCCTCGAAGCGATTTCTTTGATC 336
Db 56 ProSerGlnAspProLeuValMetTrpThrAsnGlyGlyProGlyCysSerSerLeuGly 75
QY 337 GGTCTCTTCGAGAGTGGGCGCTTCCCATGTCATTCGACTTTGATGACTACATC--- 393
Db 76 GlyGluAlaSerGluHisGlyLeuPheLeuValAsnAlaAspGlyAlaThrIleThrArg 95
QY 394 AACCTCTACTCTGGAAACGAGGTCTCCATTTACTATTCTCTCCAGCCATTTGGAGTC 453
Db 96 AsnProTyrSerTrpAsnArgValSerAsnIleLeuTyrIleGluGlnProValGlyVal 115
QY 454 GCTTTTTCATAGTATAGTGGTGTGATGGTCCATTAAACCTCTAATCGGCTCGCGAA 513
Db 116 GlyPheSerTyrSerAsnSerThrAsp----- 124
QY 514 AATTCGAGCTTTTCAGAGAGTTCAGGCGCGTATCCCAACCATTCGATGCCACTCTGATCGAT 573
Db 125 -----Asp 125
QY 574 ACTACCAATCTT-----GCCGAGAGGCGCTTGGGAGATCCTGCAAGGATTC 621
Db 126 TyrGlnAsnLeuAsnAspValGlnAlaSerAspMetAsnAlaLeuArgAspPhe 145
QY 622 CTTAGTGGACTACTAGTCTGGACTCTAGGTCAGCTCTAAGGACTTCACTCTATGACG 681
Db 146 LeuThrArgPheProGlnPheIleGlyArg-----GluThrTyrLeuAlaGly 161
QY 682 GAGAGCTATGAGGGGCACTATGTCCT---GCATTTCTTAATCATTTTACGAGCAGAAT 738
Db 162 GluSerTyrGlyGlyValTyrValProThrThrAlaTyrAsn----- 175
QY 739 GAGAGAAATCCCAACGGTACTGTATGTTTCAGCTTAATTTCACTCTCTGGGAATT 798
Db 176 -----IleValGluGlyAsnGlyLysGlyGlnProTyrValAsnLeuValGlyIle 193
QY 799 ATT-----AACGCATCATCGAC----- 816
Db 194 LeuValGlyAsnGlyValThrAspAlaGluAlaAspSerAsnSerIleProMetMet 213
QY 817 -----GAGGCGATCCAGGCCCTTACTACCTGAATTCGCTGTGACAAATACCTAC 867
Db 214 LysTyrHisSerLeuIleSerIleLysTyrTyrGluGlu----- 226

Search completed: November 21, 2003, 17:41:08
Job time : 51 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2003, 20:31:11 ; Search time 6064 Seconds
(without alignments)
11252.846 Million cell updates/sec

Title: US-09-712-338-1

Perfect score: 1668

Sequence: 1 atgcgggctagcaattctt.....ccagtggtgtagtcgcatag 1668

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.man.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1662	99.6	1662	6	AR129928 Sequence
2	993	59.5	2245	8	AF394242 Aspergill
3	894.6	53.6	1656	6	AX534871 Sequence
4	313.6	18.8	3150	6	AX534814 Sequence
5	134.4	8.1	1872	6	AX534872 Sequence
6	118.6	7.1	1665	6	AX534866 Sequence
7	116.8	7.0	1581	6	AX534824 Sequence
8	116.8	7.0	2940	6	AX534767 Sequence
9	114.6	6.9	3080	6	AX534809 Sequence
10	107.8	6.5	3221	6	AX534815 Sequence
11	86.4	5.2	2503	6	A75535 Sequence 2
12	86.4	5.2	2503	6	I70282 Sequence 3
13	84	5.0	2002	6	I33983 Sequence 3
14	84	5.0	2002	6	I74375 Sequence 3
15	84	5.0	2002	6	I77239 Sequence 3
16	82.4	4.9	1611	6	AX534826 Sequence
17	78.2	4.7	2860	6	AX534769 Sequence
18	76	4.6	2068	6	I33982 Sequence 1
19	76	4.6	2068	6	I74374 Sequence 1
20	76	4.6	2068	6	I77238 Sequence 1
21	74.6	4.5	3850	6	E12103 DNA encodin
22	74.6	4.5	3850	8	PPERCIGEN
23	72.4	4.3	860	8	CNSOLBTO
24	70.6	4.2	2214	8	PAU67174
25	68.4	4.1	1368	6	AX534850
26	68.4	4.1	1653	6	AX536918
27	68.4	4.1	2441	6	AX534793
28	68.2	4.1	4308	8	D86560
29	68.2	4.1	37000	8	SPAC19G12
30	67.6	4.1	1985	8	YSACARPEPY
31	66.2	4.0	2632	8	AR129959
32	66.2	4.0	2632	8	YSCPRCCPY
33	66.2	4.0	1581	8	SC8175
34	65.6	3.9	2280	8	AB051820
35	65.6	3.9	2509	8	AF085063
36	63.2	3.8	1446	6	AX534852
37	63.2	3.8	2520	6	AX534795
38	61.8	3.7	1527	6	AX594884
39	61.8	3.7	2016	8	SCVBR139W
40	61.8	3.7	2027	6	AX536454
41	61.8	3.7	29686	8	SC29711
42	61.8	3.7	50277	2	AC138524 5
43	59.6	3.6	1580	8	HVACXPII2
44	59.4	3.6	4760	3	MQSAACR
45	59.2	3.5	1664	3	AK115283

ALIGNMENTS

RESULT 1	AR129928	Sequence 1	1662 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	AR129928	Sequence 1 from patent US 6187578.				
DEFINITION	AR129928					
ACCESSION	AR129928					
VERSION	AR129928.1	GI:14117825				
KEYWORDS						
SOURCE		Unknown.				
ORGANISM		Unknown.				
REFERENCE		1 (bases 1 to 1662)				
AUTHORS		Blinkovsky, A., Berka, R., Rey, M., Golightly, E., Klotz, A., Mathisen, T. Erik., Danbmann, C. and Brown, K.M.				
TITLE		Carboxypeptidases and nucleic acids encoding the same				
JOURNAL		Patent: US 6187578-A 1 13-FEB-2001;				

FEATURES		Location/Qualifiers
source		1..1662
BASE COUNT		396 a 468 c 398 g 400 t
ORIGIN		
Query Match		99.6%; Score 1662; DB 6; Length 1662;
Best Local Similarity		100.0%; Pred. No. 0;
Matches 1662; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
QY	1	ATCGTGGGTACGAAATTTCTCTCAGTCTACCTTGTGACGCGAGTTGGGCGCTTCCA 60
Db	1	ATCGTGGGTACGAAATTTCTCTCAGTCTACCTTGTGACGCGAGTTGGGCGCTTCCA 60
QY	61	GGAAGTACACGGCGTCCGTGGTAGAGACAGTACCCCAAGAACCCACCGGGGTCAAG 120
Db	61	GGAAGTACACGGCGTCCGTGGTAGAGACAGTACCCCAAGAACCCACCGGGGTCAAG 120
QY	121	ACTCTTACACCGCAAAATGTCCACATCCGGTACAAAGAAACCGGGGCGAGGGCGTC 180
Db	121	ACTCTTACACCGCAAAATGTCCACATCCGGTACAAAGAAACCGGGGCGAGGGCGTC 180
QY	181	TGCAGACTACCCCGGTGTCAAAATCTCTCTGATATGCGACACCTCTCCCGAGTCC 240
Db	181	TGCAGACTACCCCGGTGTCAAAATCTCTCTGATATGCGACACCTCTCCCGAGTCC 240
QY	241	CATACCTTCTTCTGTTCTTTCGAGCCAGACATACCCAGAACTGCACTTATCACATTG 300
Db	241	CATACCTTCTTCTGTTCTTTCGAGCCAGACATACCCAGAACTGCACTTATCACATTG 300
QY	301	TGGTTGAATGTGGCGCTGGAGCGAATCTTTTGAATCGGTCTCTTCGAAGAGTTGGCGCT 360
Db	301	TGGTTGAATGTGGCGCTGGAGCGAATCTTTTGAATCGGTCTCTTCGAAGAGTTGGCGCT 360
QY	361	TGCCATGTCAATTCGACATTTTGAATCAATCAACCCCTCTCTCTGAAACGAGTCTCC 420
Db	361	TGCCATGTCAATTCGACATTTTGAATCAATCAACCCCTCTCTCTGAAACGAGTCTCC 420
QY	421	AATTTACTATTCCTGTCCAGCCATTTGGAGTCGGCTTTTATATAGTATGATACGGTTAT 480
Db	421	AATTTACTATTCCTGTCCAGCCATTTGGAGTCGGCTTTTATATAGTATGATACGGTTAT 480
QY	481	GGGTCCATTAACCTCTTAATCGGGTCTGCAAAATTCGAGCTTTGCAAGAGTTCAAGGC 540
Db	481	GGGTCCATTAACCTCTTAATCGGGTCTGCAAAATTCGAGCTTTGCAAGAGTTCAAGGC 540
QY	541	CGGTACCCCAACCAATTCGATCGATCTGATCGATCACTACCAATCTTCCCGAGAGCCGCT 600
Db	541	CGGTACCCCAACCAATTCGATCGATCTGATCGATCACTACCAATCTTCCCGAGAGCCGCT 600
QY	601	TGGGAGATCTGCAAGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGGTGCAGTCT 660
Db	601	TGGGAGATCTGCAAGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGGTGCAGTCT 660
QY	661	AAGGACTTCAGTCTATGAGCGAGAGCTATGAGGGCACTATAGTCTCTGCAATCTTCAAT 720
Db	661	AAGGACTTCAGTCTATGAGCGAGAGCTATGAGGGCACTATAGTCTCTGCAATCTTCAAT 720
QY	721	CATTTTACGAGCAGAAATGAGAAATGCAAGTGTATGTTTAAATGTTTCAAGCTTAAT 780
Db	721	CATTTTACGAGCAGAAATGAGAAATGCAAGTGTATGTTTAAATGTTTCAAGCTTAAT 780
QY	781	TTCAACTCTCTGGAAATTTATTAACGGATCATCGACGAGGGATCCAGGCCCTTACTAC 840
Db	781	TTCAACTCTCTGGAAATTTATTAACGGATCATCGACGAGGGATCCAGGCCCTTACTAC 840
QY	841	CCTGAAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACGCTCAAC 900
Db	841	CCTGAAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACGCTCAAC 900
QY	901	TACATGAAGTTTGCACCAATGCAATGTTTCCAGAGTTTATTTCCACCTGCAAA 960
Db	901	TACATGAAGTTTGCACCAATGCAATGTTTCCAGAGTTTATTTCCACCTGCAAA 960

QY	961	CAGACAAACCGCACCGCATTAAGTACTAGTACGCGCTCTGCGCGAAGCCACCAACATGTCC 1020
Db	961	CAGACAAACCGCACCGCATTAAGTACTAGTACGCGCTCTGCGCGAAGCCACCAACATGTCC 1020
QY	1021	AGGGCAATGTTGAGGGGCATACAGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db	1021	AGGGCAATGTTGAGGGGCATACAGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY	1081	CATCCATATGATGACCCGAGTTCGCGCAAGTTATTAACAATAATTTCTGGCAAGGACTCT 1140
Db	1081	CATCCATATGATGACCCGAGTTCGCGCAAGTTATTAACAATAATTTCTGGCAAGGACTCT 1140
QY	1141	GTCTATGAGCTATCGCGTCAACATCACTACACCCAGTCCAAATATGACGCTCTACTAC 1200
Db	1141	GTCTATGAGCTATCGCGTCAACATCACTACACCCAGTCCAAATATGACGCTCTACTAC 1200
QY	1201	GCTTTCCAGCAAAACAGCGGACTTTTGTCTGCGCCAACTTATCGAAGACCTCGAGGAGATC 1260
Db	1201	GCTTTCCAGCAAAACAGCGGACTTTTGTCTGCGCCAACTTATCGAAGACCTCGAGGAGATC 1260
QY	1261	CTTGCTCTCCCGTGTCTCTCTCATCTATGCGGACGCGGATTAATCTGCACTGG 1320
Db	1261	CTTGCTCTCCCGTGTCTCTCTCATCTATGCGGACGCGGATTAATCTGCACTGG 1320
QY	1321	TTGCGCGGTGAGCGGCTTCCCTCGTGGCACTACTCCCAAGCGCGCCAGTTCCGAGC 1380
Db	1321	TTGCGCGGTGAGCGGCTTCCCTCGTGGCACTACTCCCAAGCGCGCCAGTTCCGAGC 1380
QY	1381	GCAGGTACACGCGCCCTGAAAGTCAACGGGCTGAGTATGGGAACTCGCGAGTATGGT 1440
Db	1381	GCAGGTACACGCGCCCTGAAAGTCAACGGGCTGAGTATGGGAACTCGCGAGTATGGT 1440
QY	1441	AATTTCTCTTCACTCGCGTCTATGAGCGAGCCCATGAAGTCCCATATACCGAGCCCATC 1500
Db	1441	AATTTCTCTTCACTCGCGTCTATGAGCGAGCCCATGAAGTCCCATATACCGAGCCCATC 1500
QY	1501	GCCTCCCTGCAATGTTTAAACGGGACTATCTTGGTTGGGATATCGCAGAGCCGAGAG 1560
Db	1501	GCCTCCCTGCAATGTTTAAACGGGACTATCTTGGTTGGGATATCGCAGAGCCGAGAG 1560
QY	1561	AAGATCTGCGCCAGCTACAAAGCAAGTCAACGGGCTGAGTATGAGTACGATACAGTCTCC 1620
Db	1561	AAGATCTGCGCCAGCTACAAAGCAAGTCAACGGGCTGAGTATGAGTACGATACAGTCTCC 1620
QY	1621	GTGCGGCTGCTTACGGCTACGAGTATCCAGTCTCCAGTCTGCTATG 1662
Db	1621	GTGCGGCTGCTTACGGCTACGAGTATCCAGTCTCCAGTCTGCTATG 1662

RESULT 2	2245 bp	DNA	linear	PLN 24-JUL-2001
AF394242				
LOCUS	Aspergillus oryzae strain TK3	carboxypeptidase S1 (cpl) gene,		
DEFINITION	complete cds.			
ACCESSION	AF394242			
VERSION	AF394242.1	GI:15004615		
KEYWORDS				
SOURCE	Aspergillus oryzae			
ORGANISM	Aspergillus oryzae			
REFERENCE	1. (bases 1 to 2245)			
AUTHORS	van den Broek, P.			
TITLE	Direct Submission			
JOURNAL	Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,			
Lausanne	CH-1000, Switzerland			
Location/Qualifiers				
1..2245	/organism="Aspergillus oryzae"			
	/mol_type="genomic DNA"			
	/strain="TK3"			
	/db_xref="taxon:5062"			

[illegible]

Db	1681	GCACAGTTATTAACAATAATTTCTGGCCAAAGGACTCTGTCTATGAGCCTATCGCGGTCAA	1740	1	ATCGTGGCTCTCGGTTGGTGGTCTTTGTTGCCCTGGCTGCACTTACTTGTGTATGCC	60
Qy	1164	CATCAACTACACCGAGTCCAAATATGACGTCTACTAGCTTTCCAGCAACACGCGACTT	1223	61	GGAGTATACCGGCGTCCGTCGGTAGAAGACAGCTACCCCAAGAACCCACCGGGGTCAAG	120
Db	1741	CATCAACTACACCGAGTCCAAATATGACGTCTACTAGCTTTCCAGCAACACGCGACTT	1800	61	GAGAAATGAATGGTCACTCTACGATAAGAAGCGAGTTACCAAAAGCGTCCATGGCGTCAA	120
Qy	1224	TGCTGCGCCCAACTTTCATGAGACCTTCGAGGAGATCCTTGCTCTCCCGGTGGCTGTCTC	1283	121	ACTCTTACACCGCAACACATGTCACCATCGGTACAAAGAACCCCGGGGAGAGGGCGTC	180
Db	1801	TGCTGCGCCCAACTTTCATGAGACCTTCGAGGAGATCCTTGCTCTCCCGGTGGCTGTCTC	1860	121	TCGATAAAAAACCCCAACAATGTCATATCAGGTATTAAGAACCAAGAACCGAAGGAAT	180
Qy	1284	CCTCATCTATGGGACCGGCAATATCATCTGCAACTGGTTGGCGGTGAGCGCGTTCCCT	1343	181	TGCAGACTACCCCGGTGTCAAATCCTACTCTGGATATGTGACACCTCTCCGAGTCC	240
Db	1861	CCTCATCTATGGGACCGGCAATATCATCTGCAACTGGTTGGCGGTGAGCGCGTTCCCT	1920	181	TGTGAGACACACTCTGGGTCAAATCATATCTCGGATATGTGATCTTTTCCGACAGTGG	240
Qy	1344	CGTGGCAACTCTCCAGCGCGGAGTTCCGAGCGCAGGTACAGCCCTGGAAGT	1403	241	CATACCTTCTTCTGGTCTTCGAAAGCCAGACATTAACCCAGAACTGCACCTATCACATTG	300
Db	1921	CGCTGCGAACTACTCCCAAGCGCGGAGTTCCGAGCGCAGGTACAGCCCTGGAAGT	1980	241	CATACCTTCTTCTGGTCTTCGAGTCAAGCGGTGACCCCGGAAATGATCCAGTCACTTG	300
Qy	1404	CAACGGGTGAGTATGGGAACTCGCGAGTATGGTAATTTCTCTTCACTCGGCTCTA	1463	301	TGTTGAATGGTGGCCCTGGAAAGGATTTCTTTGATCGGTCTCTTCGAAAGACTTGGGCCCT	360
Db	1981	CAACGGGTGAGTATGGGAACTCGCGAGTATGGTAATTTCTCTTCACTCGGCTCTA	2040	301	TGGCTGAATGGTGGCCCTGGAAAGGATTTCTTTGATGGGCTTTTGAAGACTTGGTCCG	360
Qy	1464	TGAGCGAGGCAATGAGTCCCATATACAGCCCATCGCTCCCTGGAATTTGTTAAACG	1523	361	TGCCATGTCAATTCGACTTTTGTATGACTATACATCAACCTCACTCTGTGGAACGAGTCTCC	420
Db	2041	TGAGCGAGGCAATGAGTCCCATATACAGCCCATCGCTCCCTGGAATTTGTTAAACG	2100	361	TGTCAATCACACGAGTACGAATCAATCAATCACTAGTACTCTCTGGAACGAGTCAAC	420
Qy	1524	GACTATCTTGGTGGATATCGAGAGGCGCAGAGAGATCTCGCCCGAGTCAAGAC	1583	421	AATTTACTATTTCTGTCCAGCCATTTGGAGTGGCTTTTCATATAGTATGATACGTTGAT	480
Db	2101	GACTATCTTGGTGGATATCGAGAGGCGCAGAGAGATCTCGCCCGAGTCAAGAC	2160	421	AATCTTCTTCTTCTGTCTCAGCCCTCGGTGGGTTCTCTTACAGTGAACCGAGGCC	480
Qy	1584	GAATGGAAGGCTACAGTACGATACACAGTCGTCGGTCCGCTGCGCTACGGCTACCG	1643	481	GGGTCCATTAACCTGTAACTGGGTCTCGAAATTCGAGCTTTCGAGGAGTTTCAGGGC	540
Db	2161	GAATGGAAGGCTACAGTACGATACACAGTCGTCGGTCCGCTGCGCTACGGCTACCG	2220	481	GGGTCCATTAACCTGTAACTGGGTCTCGAAATTCGAGCTTTCGAGGAGTTTCAGGGC	540
Qy	1644	CATGTCAGTGTGGTATGGCATAG 1668		541	CGGTACCCAAACCATTTGATGCGCATCTGTATGATATCAATCAATTTTGGCGCAGAGCCCT	600
Db	2221	CATGTCAGTGTGGTATGGCATAG 2245		541	CGATACCCAGTTATTTGATGCGCATCTGTATGATATCAATCAATTTTGGCGCAGAGCCCT	600
RESULT 3	AX534871	1656 bp	DNA	linear	PAT 22-NOV-2002	
LOCUS	AX534871	Sequence 109 from Patent WO02068623.				
DEFINITION	AX534871					
ACCESSION	AX534871.1	GI:25261462				
VERSION						
SOURCE	Aspergillus niger					
KEYWORDS	Aspergillus niger					
ORGANISM	Aspergillus niger					
REFERENCE						
AUTHORS	Bukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.					
TITLE	Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A., Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W., Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiebler, J. and Albang, R.					
JOURNAL	Novel genes encoding novel proteolytic enzymes					
FEATURES	Patent: WO 02068623-A 109 06-SEP-2002; DSM N.V. (NL)					
source	Location/Qualifiers					
	1..1656					
	/organism="Aspergillus niger"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:5061"					
BASE COUNT	411 a 426 c 409 g 410 t					
ORIGIN						
Query Match	53.6%; Score 894.6; DB 6; Length 1656;					
Best Local Similarity	71.3%; Pred. No. 3.9e-221;					
Matches 1179; Conservative	0; Mismatches 474; Indels 0; Gaps 0;					
Qy	1	ATCGTGGCTCTCGGTTGGTGGTCTTTGTTGCCCTGGCTGCACTTACTTGTGTATGCC	60	1021	AGGCAATGTTGAGGGGCCATATCTAGCCCTTTGCTGGTCTGGTGTGTATGATATTCGG	1080
				1021	AGGCAATGTTGAGGGGCCATATCTAGCCCTTTGCTGGTCTGGTGTGTATGATATTCGG	1080
				1081	CATCCATATGATGACCGGACTCGGCCAGTATTATACAAATTTCTGGCAAGAGACTCT	1140
				1081	CACCCCTACAATGACCGGACCCCGCGCTCTTCTTGTGACTACCTCAAGAAAGACTCA	1140

```

QY 1141 GTCATGAGCGTATCGCGCTCAACATCACTACACCGAGTCCCAATAATGACGTCTACTAC 1200
Db 1141 GTCATGAGTGTATCGCGGTGGACATTAATCAACCGAGTCCAGCGGCGAAGTATATAT 1200
QY 1201 GCTTTCCAGAAACAGCGGACCTTTGTCTGCGCCCACTTCATCGAAGACCTTCGAGGAGATC 1260
Db 1201 GCATTCAGACAGACCGCGGACCTTTGTATGGCCGAATTTCAATTCAGGACCTCGAAGAGATC 1260
QY 1261 CTGTCTCTCCCGGTGGTGTCTCCCTCATCTATGCGGACGCCGATTAATCTCRACTGG 1320
Db 1261 CTCCAACTCCCGGTAGCGGTGCTGTATCTAGCGGATGCCGACTATATCTGTAACTGG 1320
QY 1321 TTCGGCGGTAGCGGCTTTCCCTCGCTGCGAACTACTCCCAAGCCGCCAGTTCCGAAGC 1380
Db 1321 TTCGGCGGTAGCGGCTTCTCACTCGCAGTTAACTACCCCACTGACGCTCAGTTCGCTGCA 1380
QY 1381 GCAGGGTACACGCCCTGAAAGTCAACGGGTGCGATGATGGGAAACTCGCGAGTATGTT 1440
Db 1381 GCGGGATACACACCCCATGACAGTAGATGGGTGGAATACGGTACGACTCGCGAGTATGTC 1440
QY 1441 AATTTCTCTTCACTCGCTCTATGAGGCGGCGCATGAGTCCCATACACGCGGCTATC 1500
Db 1441 AACTTTTCGTTACCGCGGTATATCAGGCTGGGCGAGGTTCCATCTATCAACCGATC 1500
QY 1501 GCTCCCTGCAATGTTTAAACCGGACTATCTTCGTTGGGTATCGCAGAGGCGCGAAG 1560
Db 1501 GCAGCGTTCGAGTGTTCACCGCTACTTATTTGGATGGATATTCGACGGGTACAACT 1560
QY 1561 AGACTCTGCCAGCTACAGAGGATGAAACGGCTACAGTACGCGATACACAGTCTGCC 1620
Db 1561 CAGATTGCGCCGAAATATAGCAACCAACGGGACATCGCAGGCTACACACGCGAGTCTGTC 1620
QY 1621 GTGCCGCTGCCTACGCGCTACCGAGTATCCAGT 1653
Db 1621 GTGCCACTGTCCAGCGGTGAGTACCGTCAAT 1653

```

```

RESULT 4
AX534814
LOCUS AX534814 3150 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 52 from Patent WO02068623.
ACCESSION AX534814
VERSION AX534814.1 GI:25261343
KEYWORDS
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
REFERENCE 1
AUTHORS Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A., Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gustedt Von,W., Heinrich,O., Maier,D., Spreafico,F., Folkers,U., Hopper,S., Kemmer,W., Tan,P., Stiebler,J. and Albang,R.
TITLE Novel genes encoding novel proteolytic enzymes
JOURNAL Patent: WO 02068623-A 06-SEP-2002;
DSM N.V. (NL)
FEATURES
source
1. 3150
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT 801 a 778 c 738 g 833 t
ORIGIN
Query Match 18.8%; Score 313.6; DB 6; Length 3150;
Best Local Similarity 72.5%; Pred. No. 3.6e-70;
Matches 406; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```

```

QY 1091 ATGACCCGACTCGGCAAGTATTAACAACAATTTCTGCGCAAGGACCTGTGATGGACG 1150
Db 2194 AGGACCCGACCCCGCGCTCTACTTTGTGTGACTACCTCAAGAAAGACTCAGTCAATGAT 2253

```

```

QY 1151 CTATCGCGCTCAACATCACTACACCGAGTCCCAATAATGACGTCTACTACGTTTCCAGC 1210
Db 2254 CTATCGCGCTGGACATTAATCAACCGAGTCCAGCGGCGAAGTATATATGATCAATTCAGC 2313
QY 1211 AAACAGCGGACCTTTGTCTGCGCCCACTTCATCGAAGACCTTCGAGGAGATCTCTTGTCTCC 1270
Db 2314 AGACCGGCGACTTTGTATGCGCGAATTTCAATTCAGGACCTTCGAAGAGATCTCTCAACTCC 2373
QY 1271 CGTGGGTGTCTCCCTCATCTATGCGGACGCCGATTAATCTGCAACTGTTTCGGGGTC 1330
Db 2374 CGGTACGCGTGTGTTGATCTACGGCGATGCCGACTATATCTGTAACCTGTTTCGGCGGTC 2433
QY 1331 AGCGCGTTTCCCTCGCTCGGAACTACTCCCAAGCCGCCAGTTCCGAAGCGGAGTACA 1390
Db 2434 AGCCCACTCTACTCGCAGTTAACTACCCCACTGACGCTCAGTTCGTCGAGCGGATACA 2493
QY 1391 CGCCCTCGAAAGTCAACCGCGCTCGAGTATGGGAAACTCGCGAGTATGTAATTTCTCCT 1450
Db 2494 CACCCATGACAGTAGATGGGTGCAATACGTTGAGACTCGGAGTATGGCACTTTTCGT 2553
QY 1451 TCACCTCGCTCTATGAGGCGGCGCATGAACTCCCATCTACACGAGCCCATCGCCTCCCTGC 1510
Db 2554 TCACCGCGCTATATCAGGCTGGGCGACGAGTTCCATCTATCAACGATCGCAGCGTGC 2613
QY 1511 AATTGTTTAAACCGGACTATCTTCGTTGGGTATCGCAGAGGCGCGAAGAGATCTGCG 1570
Db 2614 ACTGTTTAAACCGTACTTTTATTTGGATGGATATTTGACGCGGTACCACTCAGATTGGC 2673
QY 1571 CCAGCTACAAGACGAATGGAACGGTACAGCTACGCATACAGTCTCGTCCGTCGCGCTGC 1630
Db 2674 CCGAATATAGCACCAACGAGCATCGAGCTACACACGAGGCTGTTTCGTGCGCACTGT 2733
QY 1631 CTACGCGTACAGCATGTCTC 1650
Db 2734 CCACGCGCTCGAGTACCAACC 2753

```

```

RESULT 5
AX534872
LOCUS AX534872 1872 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 110 from Patent WO02068623.
ACCESSION AX534872
VERSION AX534872.1 GI:25261464
KEYWORDS
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
REFERENCE 1
AUTHORS Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A., Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gustedt Von,W., Heinrich,O., Maier,D., Spreafico,F., Folkers,U., Hopper,S., Kemmer,W., Tan,P., Stiebler,J. and Albang,R.
TITLE Novel genes encoding novel proteolytic enzymes
JOURNAL Patent: WO 02068623-A 110 06-SEP-2002;
DSM N.V. (NL)
FEATURES
source
1. 1872
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT 402 a 434 c 570 g 466 t
ORIGIN
Query Match 8.1%; Score 134.4; DB 6; Length 1872;
Best Local Similarity 48.1%; Pred. No. 1.2e-23;
Matches 644; Conservative 0; Mismatches 661; Indels 33; Gaps 8;

```

```

QY 240 CCATACCTTCTTGTGTTTCGAAAGCCAGACAT-----AACCCAGAAACTGCACCTAT 293
Db 282 CAATACCTTTTCTGGTACTTTCCTTCCGCGCATCACCACACAAATGATACATCCCCACT 341
QY 294 CACATTGTTGTAATGTTGCGCCCTCGAAGGATTTTGTGATCGGTCTCTTCGAAGATT 353

```

Db 342 CACCATCTGGATGAACGCGCGCGCGGATCTCCATGATTTGGGCTATTTCAAGAGAA 401
Qy 354 GGGCCCTTGCATGCTCAATTCGACTTTTGTATGACTACATCAACCTCAC---TCGTGAA 410
Db 402 CGGCCCATGACTGTGAATACGGACTCGAATTCACGGCCCTATATTCCTGGTCTGTGAA 461
Qy 411 CGAGGTCTCCAAATTTACTATTCCTGTCCAGCCATTCGGAGTGGGCTTTTCATATAGTA 470
Db 462 TGAGTACGTGATATGTTGTATATTTAGCAGCCGGTGCAGACGGATTTAGTATATGAT 521
Qy 471 TACGGTGTAGGGTCCATTAACCTGTATCTGACGGGTGTGCAAAATTCGAGCTTTGACG 530
Db 522 GTTCAGGAATGGGACGTTAGAT---TTGAATGAGACGTTTTCGTGGGGACGTTGCGG- 576
Qy 531 AGTTTCAGGCGGTTACCAACCATTTGATGCCACTCTGATCGATATCTACCAATCTTTCGCG 590
Db 577 -AGTCAGATGTGATGGGACGGTGAATGGGACGGTTAATGGGGAAGGCGCTTTGGGT 635
Qy 591 AGAGGCGGCTTTGGGAGATCTTCAGGATTCCTTAGTGGACTACCTAGCTTTGGACTTAG 650
Db 636 TGGGTTCAGGTTTGGTGGGTGAATCTCTGAATATGTTTCTTCTGTGACGGGAATGG 695
Qy 651 GGTGAGCTTAAGGACTTCAGTCTATGACGAGAGCTATGAGGGGACTATGCTCTGC 710
Db 696 TGGTGGTATGACAGGGTGAATATGACGAGGATCATATGGGGACGGTATGACCGGC 755
Qy 711 ATTCTCAATCATTTTACGAGCAGAAATGAGAGAAATGCCAAGGTAGTGT---TAATGG 767
Db 756 ATACAGCGGCTCTTTAGGAGATGAATGAGAGGATTCAGAGTGGGAGGTAAAGCAGCG 815
Qy 768 TGTTCAAGCTTAATTTCAATCTCTGGGAATTTAATACGGCATCATGACGAGGGGATCCA 827
Db 816 GAAGAAGATCCATTTGGATACGCTGGGCATTATCAATGGTGTGTGGATTTACTCGTCA 875
Qy 828 GGCCTCTTACTACCTGAAATTCGCTGTGAACAAATACCTACGTTACAGGCTGTCAACA 887
Db 876 GGTCCCTTCGTTCCCTGAGCAGCGGTATTAACAAATACGTTAGGATCGAGGGGAATCAAT 935
Qy 888 GACCTCTACAACTACATGAAGTTTGCCAAACCAATGCCAATGGTGTGCGAGATTTGAT 947
Db 936 CACGCTCTACGACCGGCTATGGATAGTTTGGAGCAGCTTGGCGGTGACGGATATGAT 995
Qy 948 TTCCACCTTCACAAACAGACAAACCGCAGCCATTAAGTGTGACTAGCGCTTGTGCGCGAAGC 1007
Db 996 CATCGAGTGTCCGATGCTGGGAGCTCGG---AGATCCCTCATCATCTGCGAGGAGGC 1052
Qy 1008 CACCAACATGTCAGGAGCAATGTTGAGGGGCCATCTACGCTTTGTGCTGTGCTGGTGT 1067
Db 1053 GTCCGACTACTGTTCCGGGAGATCAAGAGCCTGTATACGAATACCTCCGGGCGAGGATA 1112
Qy 1068 GTATGATATTCGGCATCCATATGATGACCCGACTCCGCCAGTTATTAACAATAATTTCT 1127
Db 1113 CTACGACATAGCGCATTTACCGCGGATGACGCTCTCGTGCCTTACTTCTGCGGTCTT 1172
Qy 1128 GGCAGAGGACTGTGTCATGAGCGCTATCGGCGTCAACATCAACTACCCAGTCCCATATA 1187
Db 1173 GAATGCCCATGGGTCAAAAGGCACTTGGGGTCCCGGTGAATATACCATGTCTGTCAGA 1232
Qy 1188 TGAAGTCTACTAGCTTTCCAGCAACAGCGGACTTTGTCTGCGCCAACTTCATCGAAGA 1247
Db 1233 GGCAGTGGGGAACAGTTTCCTCTGACGGGCGAATATCCGGAATATGCCCGCGAAT 1292
Qy 1248 CTTGAGGAGA-----TCCTTGTCTCTCCCGTGGTGTCTCCCTCATCTATGGCGA 1298
Db 1293 GATCGGGGATTTGGATACTTGTCTGACTCCGGTGTCAAGGTGGCTATGATATGGGGA 1352
Qy 1299 CGCCGATTAATCTGCAACTGTTTCGGGTGAGCGGCTTCCCTGCTCGGACTACTC 1358
Db 1353 CGGGGACTATGCTTGTTCGGTGGCGCGCGGGGAGATGTCAGCCTGCTGTGTGAGTACGA 1412
Qy 1359 CCAAGCGGCCAGTTCCGAGCGCAGGGTACAGCCCTTGAAGTCAAGCGGCTCGAGTA 1418

Db 1413 GGATCGGGAAGTTCGCTGCTGGTATCGGAAGTCAGACGAGATCATCTAGT 1472
Qy 1419 TGGGG---AAATCGCGAGTATGTAATTTCTCTTCACTCGCTCTTATGAGGAGGCCA 1475
Db 1473 TGGGGTCTAGTAAGGAGTATGGAACTTCTCGTTCAACGCTGTCTTTAGGCGGGCCA 1532
Qy 1476 TGAAGTCCCATACTACAGCCCATCGCTCCCTGCAATTTGTTAAACGGACTATCTTCGG 1535
Db 1533 TGAGGTGCCATTTTATCAGCCCGAACGGCGTATGAGATTTTATCGCGCTCAGTTAA 1592
Qy 1536 TTGGATATCGCAGAGG 1553
Db 1593 TTGGATATTGGACGG 1610

RESULT 6
AX534866 1665 bp DNA linear PAT 22-NOV-2002
LOCUS
DEFINITION Sequence 104 from Patent WO02068623.
ACCESSION AX534866
VERSION AX534866.1 GI:25261452
KEYWORDS
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

1
Edens, L., dijk Van, A. A., Krubasik, P., Albermann, K., Stock, A.,
Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,
Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,
Kemmer, W., Tan, P., Stiebner, J. and Albang, R.
Novel genes encoding novel proteolytic enzymes
Patent: WO 02068623-A 104 06-SEP-2002;
DSM N.V. (NL) Location/Qualifiers
source 1. 1665
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"

BASE COUNT 374 a 455 c 424 g 412 t
ORIGIN

Query Match 7.1%; Score 118.6; DB 6; Length 1665;
Best Local Similarity 47.2%; Pred. No. 1.5e-19;
Matches 673; Conservative 0; Mismatches 649; Indels 105; Gaps 6;

Qy 175 GCGCTCTCGAGACTACCCGGGTCTCAATCTACTCTGGATATGTGACACCTCTCC 234
Db 130 GGAATTTGGAAACTACGCGGTCTCCGATCTTATTCGGGCTATGTACACCTTCCCCC 189
Qy 235 GAGTCCCATACCTTCTTGGTCTTGAAGCCACATAACCCAGAACTGCACCTATC 294
Db 190 GCCTCAACAGCTTCTTTTGGTTTTCGAAGCCCGCAAGATCCCGAATCGGCTCTG 249
Qy 295 ACATTTGTTGTAATGGTGGCCCTGGAAGCGATTCTTTGATCGTCTCTTCAAGAGTTG 354
Db 250 GCATCTGGCTCAATGGCGGTCCGGTGGCTCGTCTCATGGGCTCTTGAAGATT 309
Qy 355 GGCCTTCGCATGTCAATTCGACTTTTGTATGAC---TACATCAACCTCCTCTGTTGAAC 411
Db 310 GTCTCTTGTTCATTCGATCAGACTCAAGACACAGACTCCTCAATCTTTGGAGTTGAAC 369
Qy 412 GAGTCTCAATTTACTATTCCTGTCGCCAGCTTGGAGTGGCTTTTATATAGTAT 471
Db 370 ATGAAGTCAATTTCTATTTCTTGTGACGAGCACTCAAGTGGCTTCTCATACGATGTC 429
Qy 472 AGGTTGATGGGTGCATTAACCTGTGTAAGTGGGTGCTCGAAATTCGAGCTTTGACGA 531
Db 430 CCAACAAATGGCATTGACAGTA----- 454
Qy 532 GTTCAGGGCGGTATCCCAACCATTTGATGCACTCTGATGATACCTACCAATCTTTCGCGCA 591
Db 455 -----ATGGGACTGCAATTCGGGCT 474

RESULT 8	AX534767	2940 bp	DNA	linear	PAT 22-NOV-2002
LOCUS	AX534767	2940 bp	DNA	linear	PAT 22-NOV-2002
DEFINITION	Sequence 5 from Patent WO02068623.				
ACCESSION	AX534767				
VERSION	AX534767.1	GI:25261241			
KEYWORDS					
SOURCE	Aspergillus niger				
ORGANISM	Aspergillus niger				
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.				
AUTHORS	1. Bden, L., dijk Van, A.A., Krubasik, P., Albrecht, K., Stock, A., Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., Gustedt Von, W., Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiebler, J. and Albarg, R.				
TITLE	Novel genes encoding novel proteolytic enzymes				
JOURNAL	Patent: WO 02068623-A 5 06-SEP-2002;				
FEATURES	DSM N.V. (NL)				
source	1. .2940				
	/organism="Aspergillus niger"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:5061"				
BASE COUNT	691 a 732 g 699 t				
ORIGIN					
Query Match	7.0%; Score 116.8; DB 6; Length 2940;				
Best Local Similarity	52.5%; Pred. No. 4.7e-19;				
Matches	334; Conservative 0; Mismatches 287; Indels 15; Gaps 3;				
Qy	968 ACCGACCGCATAGCTGACTAGCGCCTCTGCGCCGAAGCCACCAACATGTGCAGGACA 1027				
Db	1492 ACTGCGCGCCGAGGATCGACGAGATCTGCAGACTGCCAGCGATTTTTCGCGCAACG 1551				
Qy	1028 ATGTGTAGGGGCCATACTACGCCTTTGCTGCTGCTGTGTGTATGATATCGGATTCAT 1087				
Db	1552 AGGTGCAAAACGCTACGACATTTACCTCCGTCGGATGAGTATGACTTTTCGTGAAC 1611				
Qy	1088 ATGATGACCGACTCCGCCAAGTTATTACACAAATTTCTGGCAAAGGACTCTGTCA 1147				
Db	1612 CTCGCGACCGCTTCCCTTACGAGTTCTACGTTGACTACCTGAAACAAAGCGTCGTC 1671				
Qy	1148 ACGCTATCGGCGTCAACATCAACTACACCCAGCTCCATATATGACGTCTACTAGCT 1207				
Db	1672 CGGCATCGGCGCATACATCAATACAGGAGAGCAACACGCTGTTGGACTCGCCTTT 1731				
Qy	1208 AGCAAAACGGCAGCTTTGCTGTGGCC---AACTCATCGAAGACCTCGAGGATCC 1264				
Db	1732 CGTCCACCGGTGAACGACGGGCGACTCATGAAACACATCCAGGATGTGGCAAGCTCT 1791				
Qy	1265 CTCTCCCGCTGGTGTCTCCCTCATCTATGGCAGCGCGATTACATCTGCAACTGGTCT 1324				
Db	1792 AACAGGGTGCACGGTGTCTATGTACGCCGGGATGCCACTATACGCACTGCTG 1851				
Qy	1325 GGGGTGACGGCGCTTTCCCTCGCTCGACTACTCCCAAGCGGCCAGTTCGGAAGCGCAG 1384				
Db	1852 GTGGGGAACCGTGT-----CGTTGCAGGTCAAGGCGCCCACTTCAGTAGTGC 1902				
Qy	1385 GGTACACGCC---CTGAAAGTCAACGCGTCGAGTATGGGAACTCGCGAGTAGTGTA 1441				
Db	1903 GTTACACCAACATTTGCCTCGATGTAGTGTGACACACGCGCAGGTGCGCCAGCGGG 1962				
Qy	1442 ATTTCTCTTCTACTCGCGTCTATGAGCAGGCGCATGAGTCCCTACTACAGCCCA 1501				
Db	1963 AATTGCTTTGTGCGAGTGTATGAGGTGAGCATGAGTTCCCTTCTATCAACCCCTGC 2022				
Qy	1502 CCTCCTCGCAATGTTTAAACCGGACTATCTTCGTTGGATATCGCAGAGGGCCAGAGA 1561				
Db	2023 TTGCGTGTGAGATGTTTGAGCGCGTCAITGGCGGCAAGATGTGTGCGACGGAAAGAT 2082				
Qy	1562 AGATCTGGCCCGCTACAGACGATGGAACGGCTA 1597				